

Glosario

Términos esenciales para el mundo de los datos



A

Análisis de componentes principales (PCA)
Análisis de datos.
Análisis de datos en tiempo real.
Análisis de regresión.
Análisis de sentimientos.
Análisis de series de tiempo.
Análisis estadístico.
Análisis exploratorio de datos.
Análisis predictivo.
Aprendizaje automático.
Aprendizaje no supervisado.
Aprendizaje supervisado.
Árboles de decisión.
Árboles de regresión.

B

Big Data.

C

Clustering.
Correlación.
Crawling de datos.

D

Data Mining.
Data Warehousing.
Deep Learning.

E

Estadística descriptiva
Estadísticas inferenciales
Excel
Extracción de datos

F

Feature Engineering
Feature Selection

H

Herramientas de Business Intelligence (BI).
Herramientas de minería de datos.
Herramientas de visualización.

I

Inteligencia artificial.

K

K-means.

L

Lenguaje SQL.
Librerías de Python.

M

Machine Learning.
Manipulación de datos.
Modelos lineales.

N

Nube de puntos.

O

Optimización de sistemas.

P

Pandas.
PowerBI.
Procesamiento de datos.
Procesamiento de lenguaje natural (NLP).
Python.

S

Sistemas de aprendizaje automático.
Sistemas de detección de eventos.

T

Tableau.
TensorFlow.
Transformación de datos.
Técnicas de análisis de datos.
Técnicas de aprendizaje automático.
Técnicas de minería de datos.

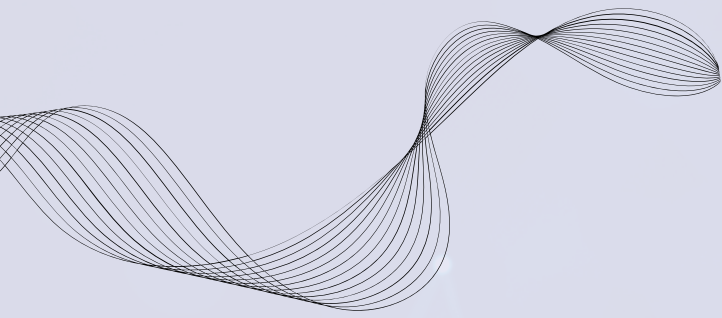
V

Visualización de datos.

W

Web Scraping.

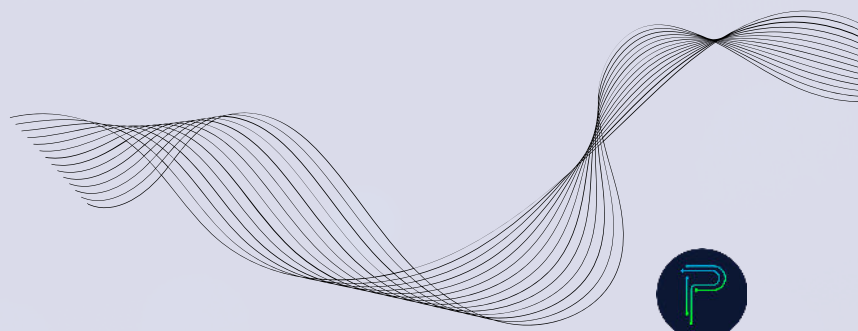




El mundo de la ciencia de datos está lleno de términos técnicos y conceptos complejos. Para poder entender y trabajar con estos datos, es esencial tener una comprensión sólida de la terminología utilizada en el campo. Como data analyst o data scientist, es importante estar familiarizado con los términos y conceptos clave utilizados en el proceso de análisis de datos.

En este glosario, se incluyen términos comunes relacionados con el proceso de análisis de datos, desde la recolección y preparación de datos hasta la visualización y la toma de decisiones. Algunos de estos términos incluyen: "Red neuronal", "Sistemas de recomendación", "Aprendizaje automático", "Minería de datos", "Pandas" y "PowerBI".

Algunos de estos términos pueden parecer complejos o confusos al principio, pero con el tiempo y la práctica, su significado y su aplicación se harán más claros. Es importante tener en cuenta que el aprendizaje nunca termina en el mundo de la ciencia de datos, ya que siempre hay nuevos desarrollos y herramientas emergentes. Sin embargo, con una buena comprensión de los términos básicos, estarás en una posición mucho mejor para entender y aprovechar estas nuevas tecnologías.





Análisis de componentes principales (PCA).

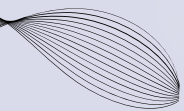
Análisis de componentes principales (PCA): El análisis de componentes principales (PCA) es una técnica estadística utilizada para analizar y visualizar datos multidimensionales. Es una forma de reducir la dimensión de los datos al identificar patrones y relaciones entre variables. El PCA busca encontrar los ejes de máxima varianza en los datos, los cuales se llaman componentes principales. Estos componentes son una combinación lineal de las variables originales y pueden ser utilizados para representar los datos de manera más comprensible.

Análisis de datos.

Análisis de datos: El análisis de datos es el proceso de examinar, limpiar, transformar y modelar datos con el fin de obtener información valiosa y tomar decisiones informadas. El análisis de datos incluye técnicas estadísticas, herramientas de visualización y algoritmos de aprendizaje automático. El objetivo es extraer información relevante de los datos y utilizarla para mejorar la toma de decisiones en una empresa u organización.

Análisis de datos en tiempo real.

El análisis de datos en tiempo real se centra en procesar y analizar los datos en el momento en que se reciben, en lugar de almacenarlos y analizarlos en un momento posterior. Se utiliza para detectar patrones y tendencias en tiempo real y para tomar decisiones en el momento. El análisis de datos en tiempo real se utiliza en una variedad de aplicaciones, como el seguimiento de vehículos, la detección de fallos en la fabricación y la monitorización de redes.



Análisis de regresión.

Análisis de regresión: El análisis de regresión es una técnica estadística utilizada para estudiar la relación entre una variable dependiente y una o más variables independientes. Se utiliza para modelar la relación entre una respuesta y un conjunto de predictores. El objetivo es encontrar un modelo matemático que pueda predecir el valor de la variable dependiente a partir de los valores de las variables independientes.

Análisis de sentimientos.

El análisis de sentimientos es el proceso de analizar y clasificar el contenido de texto para determinar el sentimiento o la opinión expresada en el mismo. Se utiliza para analizar las opiniones de los clientes sobre un producto o servicio, las tendencias en las redes sociales, y la percepción de una marca. El análisis de sentimientos se basa en técnicas de procesamiento del lenguaje natural, y puede ser realizado de forma manual o automática.

Análisis de series de tiempo.

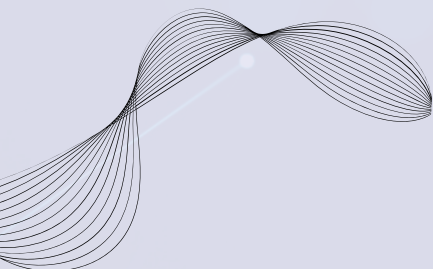
El análisis de series de tiempo es un enfoque para el análisis de datos que se centra en entender cambios en variables a lo largo del tiempo. Se utiliza para identificar patrones y tendencias en los datos a lo largo del tiempo, y para hacer predicciones sobre el comportamiento futuro. El análisis de series de tiempo se utiliza en una variedad de aplicaciones, como la economía, las finanzas, la meteorología y la medicina.

Análisis estadístico.

Análisis estadístico: El análisis estadístico es el proceso de recolectar, analizar y entender datos mediante el uso de métodos estadísticos. Incluye técnicas para recolectar datos, medir su distribución y variabilidad, y utilizar esta información para tomar decisiones informadas. El objetivo es obtener conclusiones válidas y generalizables a partir de los datos. El análisis estadístico se utiliza en una variedad de campos, desde la investigación científica hasta el negocio y la economía.

Análisis exploratorio de datos.

El análisis exploratorio de datos (AED) se centra en comprender y resumir la estructura de los datos. Se utiliza para explorar los datos, identificar patrones y tendencias, y para seleccionar los datos relevantes para un análisis más detallado. El AED se realiza antes del análisis predictivo y se utiliza para obtener una comprensión básica de los datos y para limpiar y preparar los datos para su análisis.



Análisis predictivo.

El análisis predictivo es un enfoque para el análisis de datos que se centra en utilizar los datos históricos para hacer predicciones sobre eventos futuros. Se basa en técnicas de aprendizaje automático y estadística, y se utiliza para identificar patrones y tendencias en los datos y para hacer predicciones sobre el comportamiento futuro. El análisis predictivo se utiliza en una variedad de aplicaciones, como el análisis de precios de acciones, la predicción de consumo energético y la detección de fraudes.

Aprendizaje automático.

Aprendizaje automático: El aprendizaje automático (Machine Learning, ML) es un campo de la inteligencia artificial que se ocupa de desarrollar algoritmos y técnicas que permiten a las máquinas aprender de forma automática a partir de datos, sin ser programadas explícitamente. El objetivo es permitir que la máquina realice tareas que requieren inteligencia humana, como el reconocimiento de patrones, la toma de decisiones y la resolución de problemas.

Aprendizaje no supervisado.

Aprendizaje no supervisado: El aprendizaje no supervisado es una técnica de aprendizaje automático en la que no se proporciona al modelo un conjunto de datos etiquetados para su entrenamiento. El modelo aprende a partir de los datos de forma autónoma, buscando patrones y estructuras en los datos. Un ejemplo de esto sería el uso de un conjunto de datos de transacciones bancarias para detectar patrones de comportamiento de los clientes.

Aprendizaje supervisado.

Aprendizaje supervisado: El aprendizaje supervisado es una técnica de aprendizaje automático en la que se proporciona al modelo un conjunto de datos etiquetados para su entrenamiento. El modelo aprende a partir de estos datos y se utiliza para realizar tareas de clasificación o regresión. Un ejemplo de esto sería el uso de un conjunto de imágenes etiquetadas para entrenar un modelo de reconocimiento de imágenes.

Árboles de decisión.

Un árbol de decisión es un modelo de aprendizaje automático utilizado para resolver problemas de clasificación y regresión. Es una representación gráfica de un algoritmo de toma de decisiones, donde cada nodo interno representa una prueba sobre una característica, y cada rama representa el resultado de la prueba. Los nodos hoja del árbol representan una decisión o una predicción. Los árboles de decisión son ampliamente utilizados en la clasificación de texto, la detección de fraudes y la medición de riesgos.



Árboles de regresión.

Los árboles de regresión son una variante de los árboles de decisión utilizados para resolver problemas de regresión. En lugar de predecir una clase, los árboles de regresión utilizan una serie de preguntas para predecir un valor numérico. El árbol se construye dividiendo el conjunto de datos en subgrupos más pequeños basados en las respuestas a las preguntas, y cada hoja del árbol representa una predicción numérica. Los árboles de regresión son ampliamente utilizados en aplicaciones como el análisis de precios de acciones y la predicción de consumo energético.



Big Data.

Big Data: Big Data es un término que se utiliza para describir el gran volumen de datos generados por las actividades cotidianas de las personas y las empresas. Estos datos pueden incluir información de redes sociales, transacciones comerciales, datos de sensores, entre otros. El tratamiento de estos grandes volúmenes de datos requiere tecnologías especializadas, como el procesamiento distribuido y el almacenamiento en cluster, para poder analizarlos y extraer información valiosa.





Clustering.

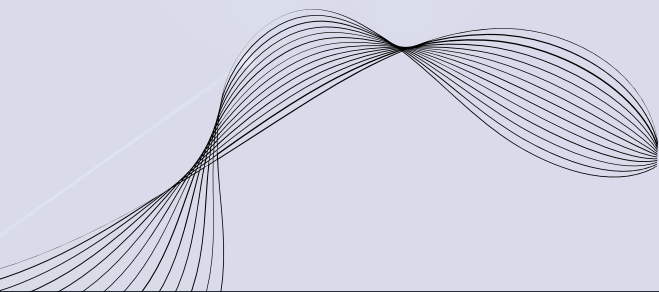
Clustering: El Clustering es una técnica de aprendizaje no supervisado que se utiliza para dividir un conjunto de datos en grupos o "clusters" de forma que los elementos de cada cluster sean similares entre sí y diferentes a los elementos de otros clusters. El objetivo es encontrar patrones y estructuras en los datos que puedan ser útiles para la toma de decisiones y la resolución de problemas. Existen diferentes algoritmos de clustering, como k-means, agrupamiento jerárquico y DBSCAN.

Correlación.

Correlación: La correlación es una medida estadística que indica la relación entre dos variables. Una correlación positiva indica que las dos variables aumentan o disminuyen juntas, mientras que una correlación negativa indica que una variable aumenta cuando la otra disminuye. El coeficiente de correlación (r) varía entre -1 y 1 , donde un valor cercano a 1 indica una alta correlación positiva, un valor cercano a -1 indica una alta correlación negativa y un valor cercano a 0 indica una baja correlación o nula.

Crawling de datos.

Crawling de datos: El crawling de datos es el proceso de recolectar automáticamente información de diferentes sitios web, utilizando programas llamados "bots" o "spiders". Estos programas navegan por la estructura de los sitios web, siguiendo enlaces y recolectando información de interés. El crawling de datos es una herramienta importante para la recolección de datos en grandes volúmenes, especialmente en el contexto del Big Data.





Data Mining.

Data Mining: El Data Mining es el proceso de extraer conocimiento y patrones valiosos de grandes volúmenes de datos. Se basa en técnicas estadísticas, de aprendizaje automático y de minería de datos para analizar y explorar los datos, con el objetivo de encontrar patrones, relaciones y tendencias que puedan ser útiles para la toma de decisiones y la resolución de problemas. El Data Mining se utiliza en una variedad de campos, como la medicina, la finanzas, el marketing y la seguridad cibernética.

Data Warehousing.

Data Warehousing: El Data Warehousing es una técnica de almacenamiento y gestión de datos que permite recopilar y almacenar grandes volúmenes de datos procedentes de diferentes fuentes, con el objetivo de facilitar su análisis y consulta. Los data warehouses se caracterizan por su gran capacidad de almacenamiento, su estructura jerárquica y su capacidad de integración de datos. Los datos almacenados en un data warehouse son históricos y están optimizados para soportar consultas complejas y análisis estadísticos.

Deep Learning.

Deep Learning: El Deep Learning es una técnica de aprendizaje automático que se basa en el uso de redes neuronales artificiales de varias capas, llamadas "redes neuronales profundas". Estas redes tienen una gran capacidad para aprender a partir de datos complejos y no estructurados, como imágenes, audio y texto. El Deep Learning se utiliza en una variedad de tareas, como el reconocimiento de imágenes, el procesamiento del lenguaje natural y la generación de contenido.





Estadística descriptiva.

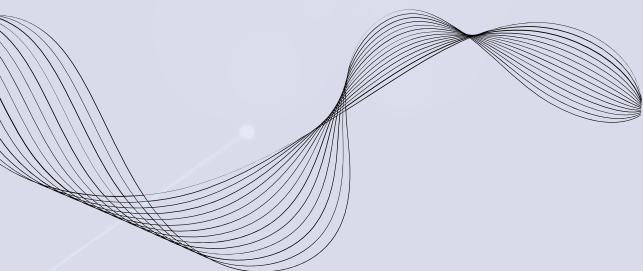
Estadística descriptiva: La estadística descriptiva es el conjunto de técnicas que se utilizan para describir, resumir y representar un conjunto de datos. Se utilizan medidas de tendencia central, como la media y la mediana, y medidas de variabilidad, como la desviación estándar y la varianza. También se utilizan gráficos y tablas para representar los datos de forma visual. La estadística descriptiva es una herramienta esencial para el análisis exploratorio de los datos.

Estadísticas inferenciales.

Estadísticas inferenciales: La estadística inferencial es el conjunto de técnicas que se utilizan para inferir conclusiones sobre una población a partir de una muestra de datos. Se basa en el uso de la probabilidad para estimar parámetros y para tomar decisiones estadísticas. La estadística inferencial se utiliza para hacer inferencias sobre la población a partir de una muestra y para hacer pruebas de hipótesis para determinar si existe una diferencia significativa entre dos grupos.

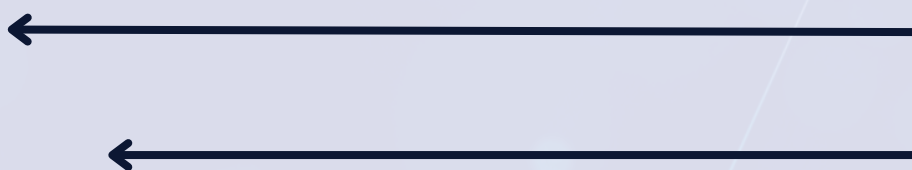
Excel.

Excel: es una herramienta de hojas de cálculo desarrollada por Microsoft. Es ampliamente utilizada para el análisis y la presentación de datos. Excel ofrece una amplia variedad de funciones para el análisis numérico y estadístico, así como herramientas de gráficos y tablas. Además, permite la conexión con bases de datos y la importación de datos de diferentes formatos.



Extracción de datos.

La extracción de datos: es el proceso de recopilar información de diferentes fuentes y almacenarla en una estructura de datos para su posterior análisis. La extracción de datos puede ser realizada manualmente o automáticamente utilizando herramientas de crawling y scraping de datos. El objetivo es obtener un conjunto de datos limpio y estructurado que pueda ser utilizado para el análisis y la toma de decisiones.



Feature Engineering.

Feature Engineering: El Feature Engineering es el proceso de crear, modificar y seleccionar las características o variables de un conjunto de datos con el objetivo de mejorar el rendimiento de un modelo de aprendizaje automático. Consiste en la aplicación de técnicas estadísticas y matemáticas para transformar y combinar las variables existentes, crear nuevas variables a partir de las existentes y seleccionar las variables más relevantes para el modelo.

Feature Selection.

Feature Selection: La Feature Selection es una técnica de preprocesamiento de los datos que se utiliza para seleccionar las características o variables más relevantes para el modelo. El objetivo es mejorar el rendimiento del modelo, reduciendo el tiempo de entrenamiento y evitando el sobreajuste. Existen diferentes métodos de selección de características, como la eliminación recursiva de características, la selección por medida de importancia y la selección por medida de información.





Herramientas de Business Intelligence (BI).

Herramientas de Business Intelligence (BI): Las herramientas de Business Intelligence (BI) son un conjunto de herramientas y tecnologías utilizadas para recopilar, almacenar, analizar y presentar información de negocios. Estas herramientas permiten a las empresas tomar decisiones informadas a partir de datos de diferentes fuentes, como ventas, finanzas, producción, entre otros. Las herramientas de BI incluyen herramientas de minería de datos, tableros de control, informes y análisis predictivo.

Herramientas de minería de datos.

Herramientas de minería de datos: Las herramientas de minería de datos son un conjunto de herramientas y tecnologías utilizadas para analizar y extraer información valiosa de grandes volúmenes de datos. Estas herramientas permiten a las empresas tomar decisiones informadas a partir de los datos, identificando patrones, tendencias y relaciones que pueden ser útiles para la toma de decisiones. Algunos ejemplos de herramientas de minería de datos son RapidMiner, KNIME y Orange.

Herramientas de visualización.

Herramientas de visualización: Las herramientas de visualización son un conjunto de herramientas y tecnologías utilizadas para representar visualmente los datos. Estas herramientas permiten a los usuarios visualizar y explorar los datos de forma intuitiva, facilitando la comprensión y el análisis de los mismos. Las herramientas de visualización incluyen gráficos, mapas, diagramas, entre otros. Algunos ejemplos de herramientas de data visualización son Tableau, Power BI, y D3.js





Inteligencia artificial.

Inteligencia artificial: La inteligencia artificial (IA) es el campo de la informática que se ocupa del estudio y desarrollo de algoritmos y sistemas que permiten a las máquinas aprender de forma automática a partir de datos. El objetivo es permitir que la máquina realice tareas que requieren inteligencia humana, como el reconocimiento de patrones, la toma de decisiones y la resolución de problemas. La IA se basa en técnicas de aprendizaje automático, procesamiento del lenguaje natural, visión por computadora y robótica.



K-means.

K-means: K-means es un algoritmo de clustering que se utiliza para dividir un conjunto de datos en k clusters. El algoritmo se basa en la asignación de cada punto del conjunto de datos al cluster cuyo centro está más cerca. El objetivo es minimizar la suma de las distancias entre los puntos y el centro del cluster al que pertenecen. El algoritmo se inicia con k centros iniciales y se repite hasta que los centros de los clusters no cambian. Es uno de los algoritmos más utilizados en clustering y es relativamente fácil de implementar y comprender.



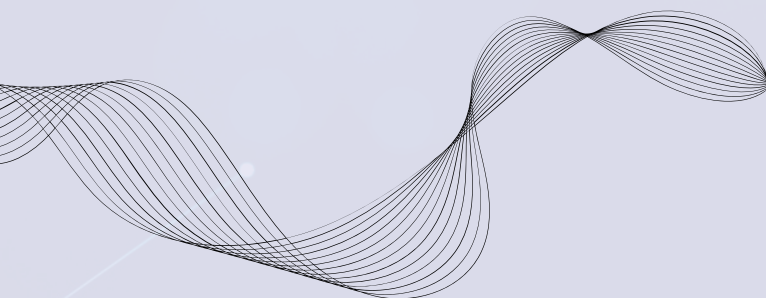


Lenguaje SQL.

LenguajeSQL (Structured Query Language): es un lenguaje de programación diseñado para trabajar con bases de datos relacionales. Es utilizado para insertar, actualizar, recuperar y eliminar datos almacenados en tablas de una base de datos. SQL permite a los usuarios escribir consultas complejas para obtener información específica de las bases de datos, utilizando comandos como SELECT, INSERT, UPDATE y DELETE. Es ampliamente utilizado en la creación y gestión de bases de datos en aplicaciones empresariales, científicas y gubernamentales.

Librerías de Python.

Las librerías de Python: son conjuntos de funciones y módulos que se pueden importar en un programa para ampliar las funcionalidades del lenguaje. Estas librerías están diseñadas para facilitar tareas específicas, como el manejo de datos, la visualización, el aprendizaje automático, entre otros. Algunos ejemplos de librerías de Python populares incluyen NumPy, Matplotlib, Scikit-learn, y NLTK.





Machine Learning.

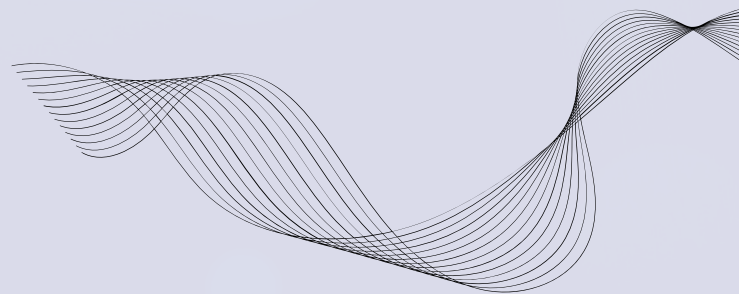
Machine Learning: El aprendizaje automático (Machine Learning, ML) es un campo de la inteligencia artificial que se ocupa de desarrollar algoritmos y técnicas que permiten a las máquinas aprender de forma automática a partir de datos, sin ser programadas explícitamente. El objetivo es permitir que la máquina realice tareas que requieren inteligencia humana, como el reconocimiento de patrones, la toma de decisiones y la resolución de problemas. Existen diferentes tipos de aprendizaje automático, como el aprendizaje supervisado, no supervisado y profundo.

Manipulación de datos.

La manipulación de datos es el proceso de limpiar, transformar, y preparar los datos para su análisis e interpretación. Esta tarea es crítica para garantizar que los datos sean compatibles con los algoritmos y técnicas de análisis utilizadas, y para obtener resultados precisos y significativos. La manipulación de datos puede incluir tareas como la eliminación de valores faltantes, la normalización de los datos, la codificación de variables categóricas y la agregación de datos.

Modelos lineales.

Modelos lineales: Los modelos lineales son una clase de modelos matemáticos que se utilizan para describir relaciones lineales entre variables. Estos modelos asumen que la relación entre las variables es una recta y se pueden utilizar para hacer predicciones o para describir la relación entre variables. Los modelos lineales se utilizan en una variedad de campos, como la economía, las finanzas, la ingeniería y las ciencias sociales. Los modelos lineales más conocidos son la Regresión lineal y el Análisis de varianza.





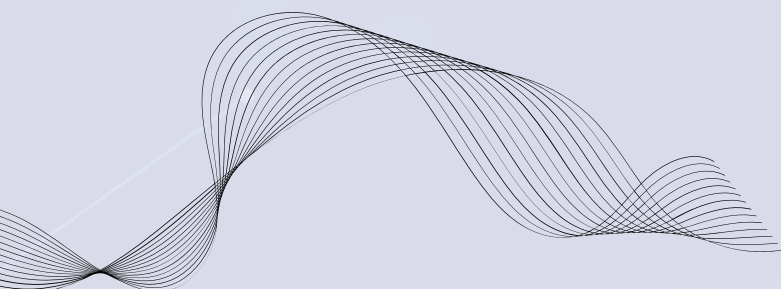
Nube de puntos.

Nube de puntos: La nube de puntos es una técnica de visualización de datos que se utiliza para representar tres o más variables de un conjunto de datos. Cada punto de la nube de puntos representa una observación del conjunto de datos y su posición en el espacio tridimensional se determina por los valores de las tres variables. La nube de puntos permite visualizar patrones y relaciones en los datos que pueden ser difíciles de ver en otras representaciones.



Optimización de sistemas.

Optimización de sistemas: La optimización de sistemas es el proceso de encontrar la mejor solución para un problema dado, maximizando o minimizando una función objetivo. La optimización de sistemas se utiliza en una variedad de campos, como la ingeniería, la economía, las finanzas y la ciencia de la computación. Existen diferentes enfoques y algoritmos para la optimización de sistemas, como la optimización matemática, la optimización heurística y la optimización global.





Pandas.

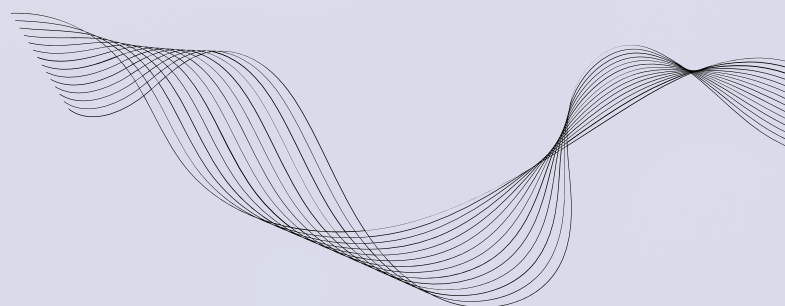
Pandas es una librería de Python para el análisis de datos. Proporciona estructuras de datos y operaciones para manipular y analizar datos de manera eficiente. Entre las principales características de Pandas se encuentra la capacidad de manejar datos faltantes, realizar operaciones estadísticas y de álgebra lineal, y realizar agrupaciones y uniones de datos.

PowerBI.

Power BI es una herramienta de visualización de datos desarrollada por Microsoft. Es utilizada para crear informes y dashboards interactivos a partir de datos de diferentes fuentes. Power BI ofrece una variedad de opciones de visualización, como gráficos, tablas, y mapas, y permite la conexión con diferentes fuentes de datos, como Excel, SQL Server, y bases de datos en la nube. Además, ofrece características avanzadas como el análisis predictivo y la integración con otras herramientas de Microsoft como Azure y Office 365.

Procesamiento de datos.

Procesamiento de datos: El procesamiento de datos es el conjunto de técnicas y procesos utilizados para transformar y limpiar los datos con el objetivo de hacerlos útiles para el análisis y la toma de decisiones. El procesamiento de datos incluye tareas como la extracción de datos, la limpieza de datos, la integración de datos, la transformación de datos y la validación de datos. El procesamiento de datos es esencial para obtener un conjunto de datos limpio y estructurado que pueda ser utilizado para el análisis y la toma de decisiones.



Procesamiento de lenguaje natural (NLP).

Procesamiento de lenguaje natural (NLP): El procesamiento del lenguaje natural (NLP) es un campo de la inteligencia artificial que se ocupa del estudio y desarrollo de algoritmos y técnicas para el procesamiento automático del lenguaje natural. El objetivo es permitir que las máquinas comprendan, generen y comuniquen en lenguaje natural de forma efectiva. El NLP se basa en técnicas de lingüística, inteligencia artificial, procesamiento de lenguaje y estadística.

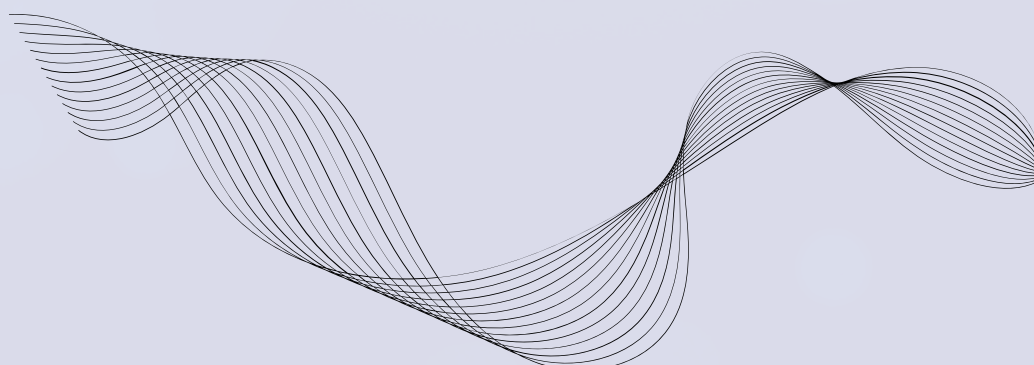
Python.

Python es un lenguaje de programación de alto nivel, con una sintaxis clara y legible. Es ampliamente utilizado en la ciencia de datos, el aprendizaje automático y el desarrollo web. Python cuenta con una gran cantidad de librerías y paquetes especializados que facilitan el trabajo con datos, como NumPy, Pandas, Matplotlib y Scikit-learn. Además, es un lenguaje interpretado, lo que lo hace fácil de depurar y desarrollar.



Sistemas de aprendizaje automático.

Los sistemas de aprendizaje automático (SAM): son una clase de algoritmos y técnicas que permiten a las máquinas aprender a partir de datos sin ser programadas explícitamente. Los SAM se utilizan en una amplia variedad de aplicaciones, como el reconocimiento de patrones, la toma de decisiones, la detección de fraudes y la mejora de la eficiencia de los procesos. Los SAM se dividen en tres categorías: aprendizaje supervisado, no supervisado y reforzado.



Sistemas de detección de eventos.

Los sistemas de detección de eventos: son un conjunto de herramientas y algoritmos que ayudan a detectar patrones y anomalías en grandes cantidades de datos. Estos sistemas se utilizan para detectar eventos relevantes en una variedad de campos, como la seguridad cibernética, la inteligencia militar, la telemetría y la detección de fallos en la fabricación.

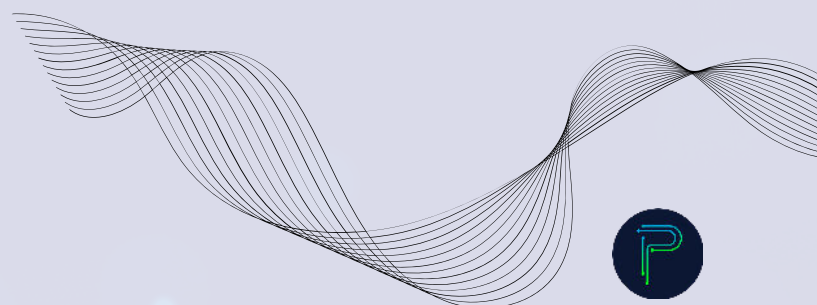


Tableau.

Tableau: es una herramienta de visualización de datos desarrollada por Tableau Software. Es utilizada para crear informes y dashboards interactivos a partir de datos de diferentes fuentes. Tableau ofrece una variedad de opciones de visualización, como gráficos, tablas, y mapas, y permite la conexión con diferentes fuentes de datos, como Excel, SQL Server y bases de datos en la nube. Además, Tableau ofrece características avanzadas como el análisis geográfico y el análisis de series de tiempo.

TensorFlow.

TensorFlow es una plataforma de código abierto para el aprendizaje automático desarrollada por Google. Se utiliza para crear y entrenar modelos de aprendizaje automático a través de una variedad de técnicas, como el aprendizaje profundo y el aprendizaje por refuerzo. TensorFlow también proporciona herramientas para la visualización de datos, el debug y el despliegue de modelos en una variedad de plataformas.



Transformación de datos.

La transformación de datos es el proceso de modificar, limpiar y preparar los datos para su análisis e interpretación. Puede incluir tareas como la eliminación de valores faltantes, la normalización de los datos, la codificación de variables categóricas y la agregación de datos. La transformación de datos es un paso importante en el proceso de análisis de datos ya que permite que los datos sean compatibles con los algoritmos y técnicas de análisis utilizadas.

Técnicas de análisis de datos.

Las técnicas de análisis de datos son un conjunto de métodos y herramientas utilizadas para explorar, limpiar, transformar y modelar datos con el fin de obtener información valiosa y útil. Los analistas de datos utilizan estas técnicas para obtener insights a partir de los datos y tomar decisiones informadas. Las técnicas de análisis de datos incluyen estadística, minería de datos, visualización de datos y aprendizaje automático.

Técnicas de aprendizaje automático.

Las técnicas de aprendizaje automático (TAA) son un conjunto de algoritmos y metodologías que permiten a las computadoras aprender a partir de datos y realizar tareas sin ser programadas explícitamente. Algunas de las técnicas más comunes de aprendizaje automático incluyen el aprendizaje supervisado, no supervisado y reforzado, el aprendizaje profundo, el aprendizaje por refuerzo y la optimización bayesiana.

Técnicas de minería de datos.

La minería de datos es el proceso de extraer información valiosa de grandes conjuntos de datos. Se utiliza para identificar patrones y tendencias que pueden ser utilizadas para mejorar los procesos de negocio, tomar decisiones informadas y generar nuevos conocimientos. Las técnicas de minería de datos incluyen análisis estadístico, aprendizaje automático, algoritmos de agrupamiento y algoritmos de reglas de asociación.





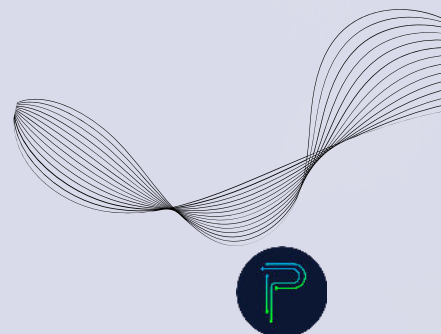
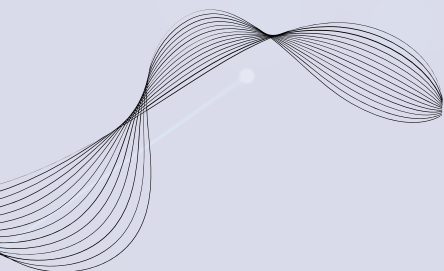
Visualización de datos.

La visualización de datos: es el proceso de representar información mediante gráficos y otros elementos visuales. Se utiliza para comunicar información compleja de manera clara y concisa, y para explorar y analizar los datos de una manera intuitiva. La visualización de datos incluye una variedad de técnicas, como los gráficos de barras, los gráficos de línea, los mapas y los diagramas de dispersión.



Web Scraping.

Web Scraping: es un técnica utilizada para extraer información automáticamente de sitios web. Es un proceso mediante el cual se automatiza la recolección de datos de una página web, utilizando scripts y programas para analizar el HTML y extraer la información deseada. Es ampliamente utilizado para la recolección de datos en investigaciones, análisis de mercado, y para la monitorización de precios y opiniones en línea.



Este glosario es solo un pequeño vistazo a la terminología utilizada en el mundo de la ciencia de datos. Sin embargo, para convertirte en un verdadero experto en el campo, es necesario obtener una formación más completa. Si estás interesado en profundizar en su comprensión de los conceptos y herramientas de análisis de datos, te recomendamos que consideres nuestro programa de máster en data science o data analytics. Ofrecemos una educación 100% práctica donde aplicarás todo lo aprendido en proyectos reales.



¡Inscríbete hoy y comienza tu camino hacia convertirte en un experto analista o científico de datos!





¿TE GUSTARÍA APRENDER MÁS SOBRE DATOS?

MÁSTER EN DATA ANALYTICS

Aprenderás Matemáticas y Estadística, Python, SQL, Machine learning y Deep learning, Visualización de datos con Tableau y Big Data para convertirte un perfil clave en cualquier empresa tecnológica.

MÁSTER EN DATA SCIENCE

Te formarás en Matemáticas y Estadística, Python, SQL, Excel, Power Bi, Visualización de datos y Big Data para convertirte un perfil clave en la toma de decisiones de cualquier empresa tecnológica.

www.pontia.tech

PONTIA.TECH

