

Módulo: Introducción al Análisis de Datos y Álgebra Lineal con Python

DATA
ANALÍTICA





NOVATO en
ML & Data Science

`lm(y~...); predic()`

Data scaling
Outliers analysis

Missing data analysis

Visualización de datos

Reglas del Juego

**Mantener el micrófono
apagado en caso de que
no vayan a hablar.**



**Nos encantaría verte.
Ten tu cámara encendida y
conozcámonos
virtualmente.**



**Preguntar en caso que
tengan dudas.**



**Disfruta de este espacio.
Desconecta del resto y
participa.**



**Por cada clase tendremos
10 min o 15 min de receso.**



Agenda

1. ¿Qué es el análisis de datos?
2. ¿Por qué necesitamos saber algebra lineal?
3. Proceso de análisis de datos
4. Niveles de análisis de datos
5. Diferentes roles en el mundo de datos
6. Introducción a Numpy

Introducción al análisis de datos con Python

¿Qué es el análisis de datos?

El análisis de datos es la práctica de trabajar con datos para obtener información útil, que luego se puede utilizar para tomar decisiones informadas.



Ejemplos en el análisis de datos



Redes Sociales

Estudio de la demografía de seguidores en Instagram para una marca o influencer.

Identificación de patrones en la actividad en línea de usuarios en Facebook para detectar características que ayudar a la compra de productos.

Ejemplos en el análisis de datos



Compañías de telecomunicaciones

Análisis de la calidad de la señal y la cobertura en una zona determinada para optimizar la red.

Análisis de las llamadas perdidas y los tiempos de respuesta en el call center para mejorar la eficiencia.

Identificación de patrones en el uso de datos de los clientes para personalizar las ofertas y mejorar la satisfacción del cliente.

Ejemplos en el análisis de datos



Monitoreo del clima

Estudio de la variabilidad climática y su impacto en la agricultura para optimizar los cultivos y prevenir pérdidas.

Análisis de la relación entre el clima y la salud para identificar factores de riesgo y prevenir enfermedades.

Estudio de la variabilidad climática y su impacto en la industria turística para optimizar la planificación y mejorar la experiencia del turista.

Ejemplos en el análisis de datos



Deportes

Análisis de la táctica y el desempeño individual de los jugadores en partidos de fútbol para mejorar el rendimiento del equipo.

Estudio de la relación entre la dieta, el entrenamiento y el rendimiento de atletas para optimizar la preparación física.

Estudio de la relación entre la forma física y el rendimiento de los corredores de maratón para optimizar la preparación.



Diferentes roles en el mundo de datos

- Analista de datos



Analista de datos en Perú
1797 resultados

Crear alerta ☐


Hace 1 semana

**Analista de Datos**
CSTI Corp
Lima, Lima, Perú (Presencial)
 Tu perfil se ajusta a este empleo

Hace 4 semanas

**Data Analyst**
Pacífico Seguros
Lima, Lima, Perú (Presencial)
 3 contactos trabajan aquí

Hace 1 mes

**Analista de Datos - Rimac -Consumo Masivo**
S&C SOLUCIONES INTEGRALES SAC
Lima, Lima, Perú (Presencial)

**Analista de datos Junior**
Vooxell
Lima, Lima, Perú (Híbrido)
 3 antiguos alumnos trabajan aquí

Hace 19 horas

**Analista de datos GCP**
Vooxell
Lima, Lima, Perú (Híbrido)
 Tu perfil se ajusta a este empleo

Hace 1 semana • 23 solicitudes

**Analista de Datos Tableau**
Vooxell
Lima, Lima, Perú (En remoto)
 Tu perfil se ajusta a este empleo

Analista de datos:

<https://www.linkedin.com/jobs/search/?currentJobId=3450199093&geoid=102927786&keywords=Analista%20de%20datos%20&location=Per%C3%BA&refresh=true>

Diferentes roles en el mundo de datos

- Analista de datos

Requisitos

- Profesionales en las especialidades de Ingeniería de Sistemas, Industrial y/o Estadística.
- Experiencia mínima de 2 años en Data.
- Indispensable manejo de SQL y Power BI.
- Manejo del idioma inglés a nivel intermedio.
- MS Office (Excel) a nivel intermedio.
- Deseable manejo de Azure Cloud.
- Deseable experiencia en venta digital.

Diferentes roles en el mundo de datos

- Analista de datos

Conocimientos

- Modelado de Datos e Inteligencia de Negocios
- Base de Datos: Oracle/SQL Server.
- Construcción y mantenimiento de ETL's.
- Lenguajes de programación: Python o R.
- Técnicas estadísticas para análisis de datos.
- Servicios de Cloud AWS: RedShift, S3, Glue, Athena
- Servicios de Cloud GCP: Bigquery, Cloud Storage
- Visualización: Power BI, Tableau, Datastudio.
- Herramientas de perfilamiento (QA).
- Gobierno de Datos - Calidad de Datos.

Diferentes roles en el mundo de datos

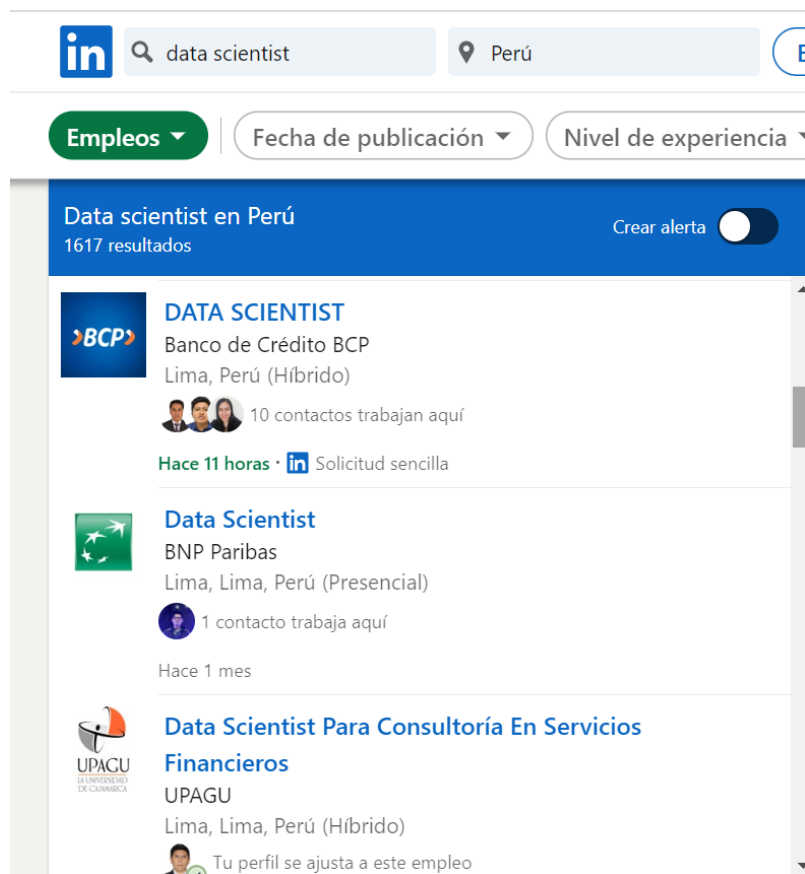
- Analista de datos

Requisitos

- Egresado de las carreras universitarias de Ingeniería de Software, Ingeniería de Sistemas o afines.
- 1 a 2 años de experiencia de implementación en soluciones de big data, analítica avanzada o afines como analista.
- Experiencia en Tableau, Power BI, Qlik.
- Conocimientos en Base de Datos Oracle SQL, MySQL y PL/SQL a nivel avanzado.
- Conocimientos de base de datos relacional y No Relacional (Elastic Search, Cassandra, Mongo DB)
- Conocimiento básico en Linux (deseable)
- Nivel de Inglés Intermedio-Avanzado (indispensable, la entrevista será en inglés).

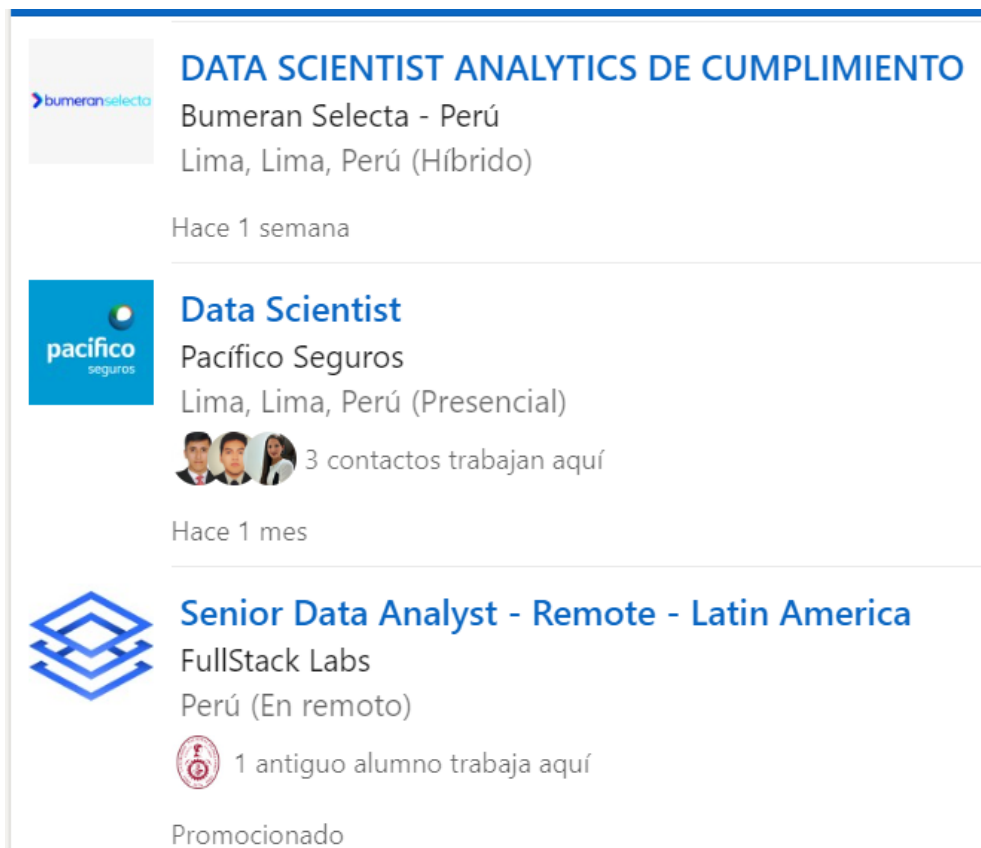
Diferentes roles en el mundo de datos

- Científico de datos



The screenshot shows the LinkedIn search interface for 'data scientist' in Peru. The search bar at the top contains 'data scientist' and 'Perú'. Below the search bar, there are filters for 'Empleos', 'Fecha de publicación', and 'Nivel de experiencia'. The results section is titled 'Data scientist en Perú' with '1617 resultados'. A toggle for 'Crear alerta' is visible. The first three results are:

- DATA SCIENTIST** at Banco de Crédito BCP, Lima, Perú (Híbrido). 10 contactos trabajan aquí. Publicado hace 11 horas. Solicitud sencilla.
- Data Scientist** at BNP Paribas, Lima, Lima, Perú (Presencial). 1 contacto trabaja aquí. Publicado hace 1 mes.
- Data Scientist Para Consultoría En Servicios Financieros** at UPAGU, Lima, Lima, Perú (Híbrido). Tu perfil se ajusta a este empleo.



The screenshot shows the LinkedIn search interface for 'Data Scientist' in Peru. The search bar at the top contains 'Data Scientist' and 'Perú'. Below the search bar, there are filters for 'Empleos', 'Fecha de publicación', and 'Nivel de experiencia'. The results section is titled 'Data Scientist en Perú' with '1617 resultados'. A toggle for 'Crear alerta' is visible. The first three results are:

- DATA SCIENTIST ANALYTICS DE CUMPLIMIENTO** at Bumeran Selecta - Perú, Lima, Lima, Perú (Híbrido). Publicado hace 1 semana.
- Data Scientist** at Pacífico Seguros, Lima, Lima, Perú (Presencial). 3 contactos trabajan aquí. Publicado hace 1 mes.
- Senior Data Analyst - Remote - Latin America** at FullStack Labs, Perú (En remoto). 1 antiguo alumno trabaja aquí. Promocionado.

Científico de datos

: <https://www.linkedin.com/jobs/search/?currentJobId=3427003006&geoid=102927786&keywords=data%20scientist&location=Per%C3%BA&refresh=true>

Diferentes roles en el mundo de datos

- Científico de datos

Perfil Técnico

- Titulado en ingeniería matemática, ingeniería de sistemas, estadística.
- 1 a 2 años participando en proyectos de ciencias de datos.
- Experiencia en Python
- Conocimiento de las herramientas y métodos de procesamiento de datos
- Construye un Scoring
- Utiliza herramientas de modelamiento estadístico
- Dominio de las A / B Testing y las herramientas / protocolos de prueba multivariante
- Utiliza herramientas de desarrollo de IT
- Dominar la visualización de datos

Diferentes roles en el mundo de datos

- Científico de datos

Requisitos

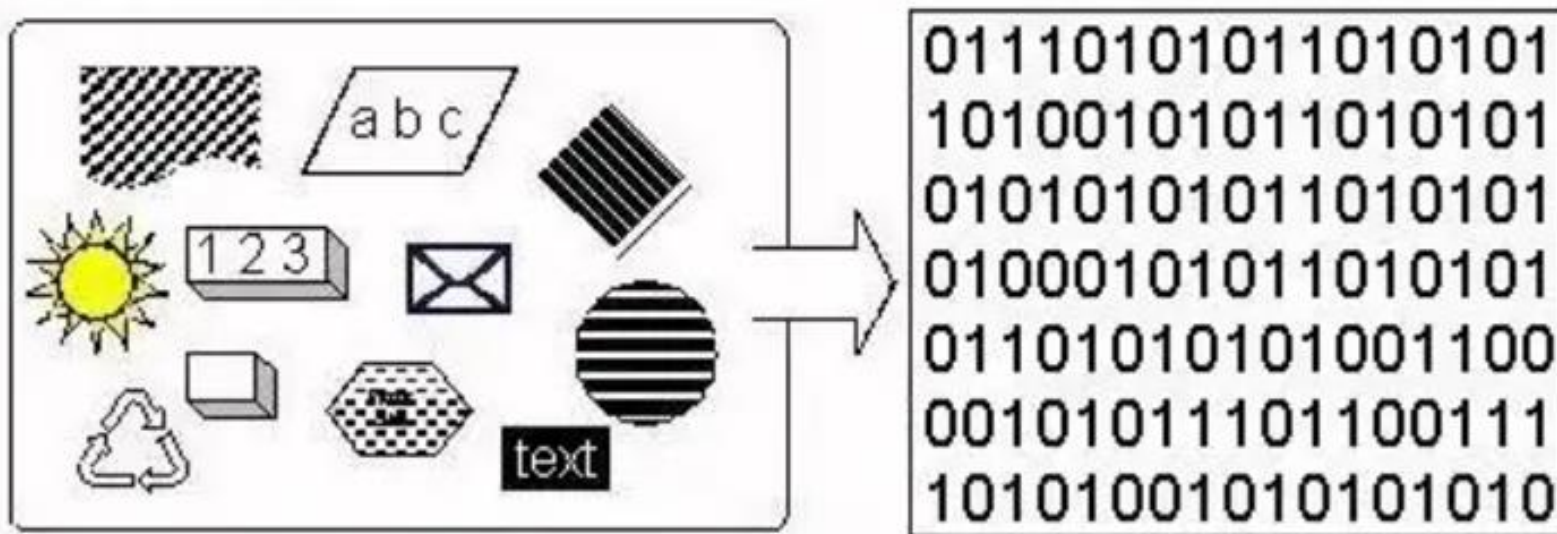
- Egresado/a, Bachiller, Titulado/a de Estadística o Ingeniería de Sistemas, Informática, Industrial, o afines con especialidad en Analytics y/o Ciencia de Datos.
- 3 años de experiencia como mínimo realizando funciones similares en empresas de servicios financieros, alta tecnología o telecomunicaciones.
- Conocimientos de lenguajes de programación analíticos Python o R, además de lograr escribir consultas SQL complejas
- Conocimientos de aprendizaje automático y estadística para resolver problemas de agrupación, clasificación, regresión, detección de anomalías, simulación y optimización en datos a gran escala.
- Experiencia de trabajo en plataformas de IA en la nube (Azure, GCP, AWS, IBM Cloud, o similares). Proceso de modelamiento end to end.
- Experiencia en el apoyo a la operacionalización, la supervisión, el mantenimiento y la mejora de modelos deseados.

Diferentes roles en el mundo de datos

- Ingeniero de datos
- Arquitecto de datos
- Analytics translator
- Ingeniero de Machine learning

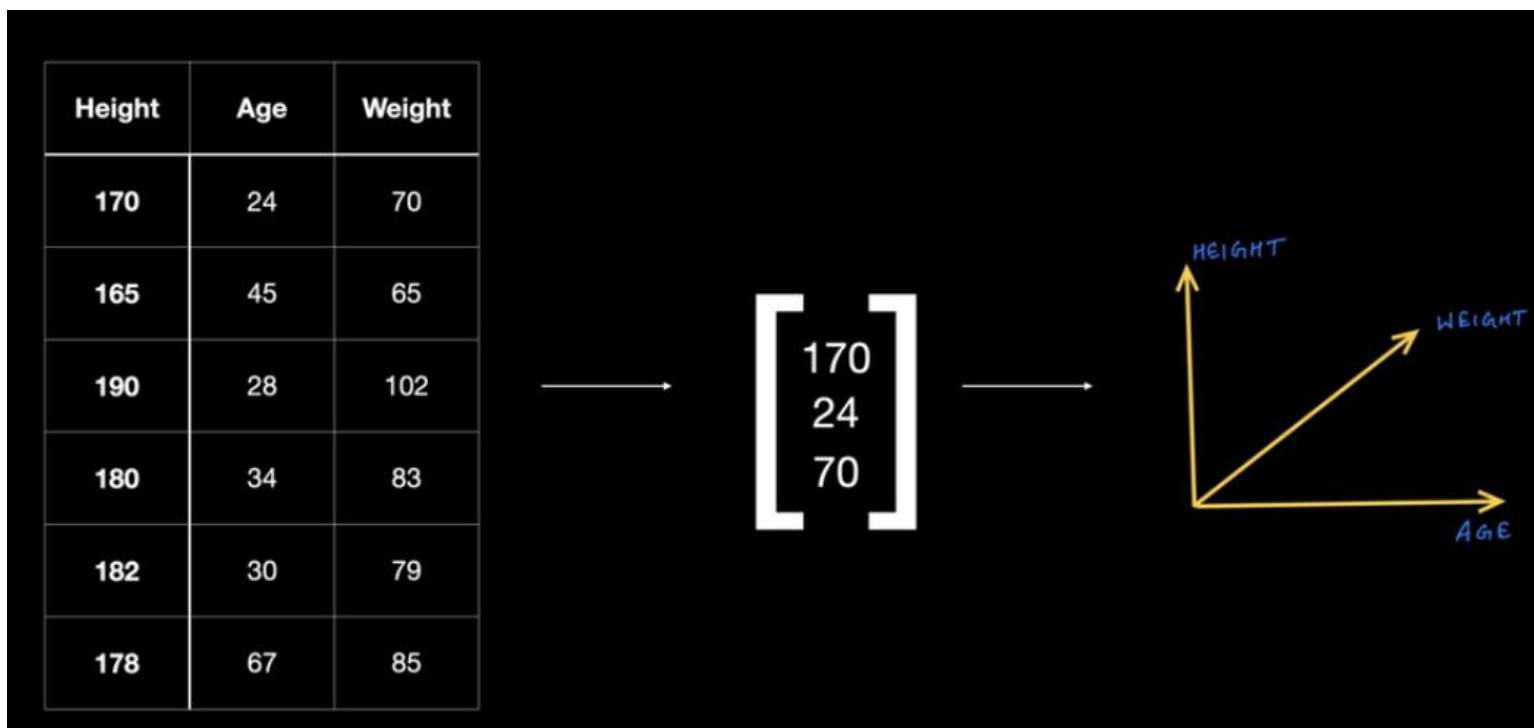
¿Por qué necesitamos saber algebra lineal?

Las máquinas sólo entienden de números. Y estos números deben representarse y procesarse de una manera que permita a las máquinas resolver problemas aprendiendo de los datos



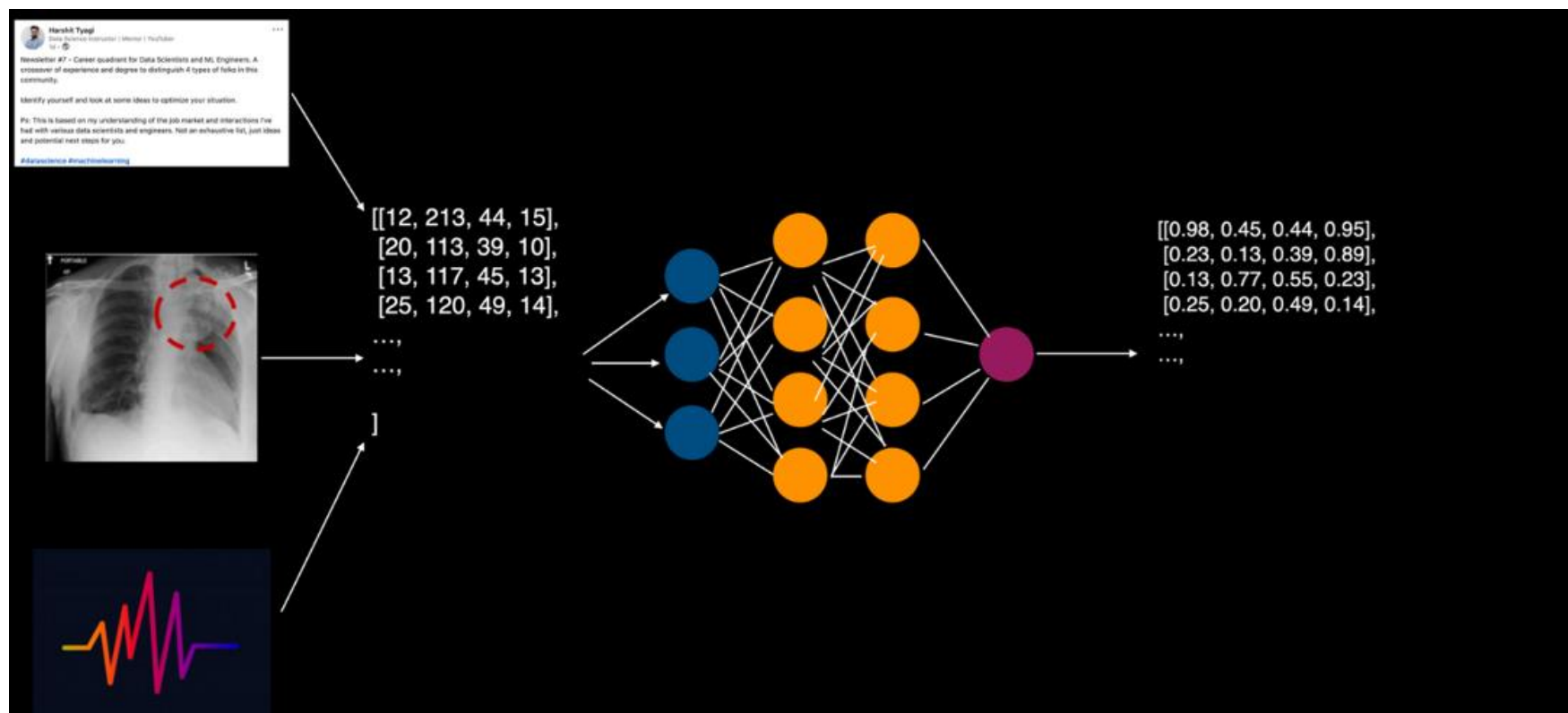
¿Por qué necesitamos saber álgebra lineal?

La mayoría de los modelos de aprendizaje automático se pueden expresar en forma de matriz



¿Por qué necesitamos saber algebra lineal?

Todo se reduce a representar en una matriz



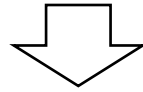
Algunas aplicaciones del álgebra lineal en la ciencia de datos

- Representación de conjuntos de datos
- Regresión
- El análisis de componentes principales (PCA)
- Aprendizaje profundo
- Sistema de recomendación

Proceso de análisis de datos

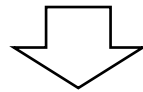
Identificación

¿Qué problema está tratando de resolver la empresa? ¿Qué necesita medir y cómo lo medirá?



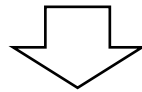
Recopilación de datos

Puede provenir de fuentes internas o de fuentes secundarias.

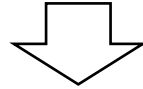


Limpieza de datos

Depuración de datos duplicados y anómalos, la resolución de incoherencias, la estandarización de la estructura y el formato de los datos y el tratamiento de los espacios en blanco y otros errores de sintaxis.

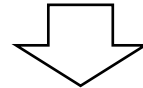


Proceso de análisis de datos



Análisis de datos

Hacemos uso de diferentes técnicas y herramientas de análisis de datos



Interpretación de resultados

¿Qué recomendaciones puede hacer con base en los datos? ¿Cuáles son las limitaciones de sus conclusiones?

Niveles de análisis que necesita para una mejor toma de decisiones

Análisis descriptivo

¿Qué sucedió?

Análisis de diagnóstico

¿Por qué sucedió?

Análisis predictivo

¿Qué podría pasar en el futuro?

Análisis prescriptivo

¿Qué acción tomar?

Introducción a



¿Qué es Numpy?

Numpy (Numerical Python)

- Biblioteca Python de código abierto
- Es el estándar universal para trabajar con datos numéricos en Python
- Contiene matrices multidimensionales y estructuras de datos matriciales
- Proporciona ndarray, un objeto de matriz n-dimensional homogéneo, con métodos para operar de manera eficiente en él.
- Se puede utilizar para realizar una amplia variedad de operaciones matemáticas en matrices.

Reto de clase:

Resuelve los 10 ejercicios de Numpy usando lo aprendido durante la clase de introducción a Pandas

Enviarlo al final de la sesión por correo con el asunto:
Ejercicios Numpy - Módulo Introducción al Álgebra Lineal
- [Apellidos y nombres]

Correo: team@dataanalitica.net

Challenge 3:

Elaborar 5 ejercicios (Python) de lo aprendido durante la clase de introducción a NumPy

Enviarlo por correo con el asunto: Challenge 3 – Módulo Introducción al Álgebra Lineal – [Apellidos y nombres]

Correo: team@dataanalitica.net

Challenge 4:

Resolver los ejercicios (Python) de lo aprendido durante la clase de introducción al álgebra lineal

Enviarlos por correo con el asunto: Challenge 4 – Módulo Introducción al Álgebra Lineal – [Apellidos y nombres]

Correo: team@dataanalitica.net