



Programació
amb Python



Unió Europea
Fons social europeu

Data Science

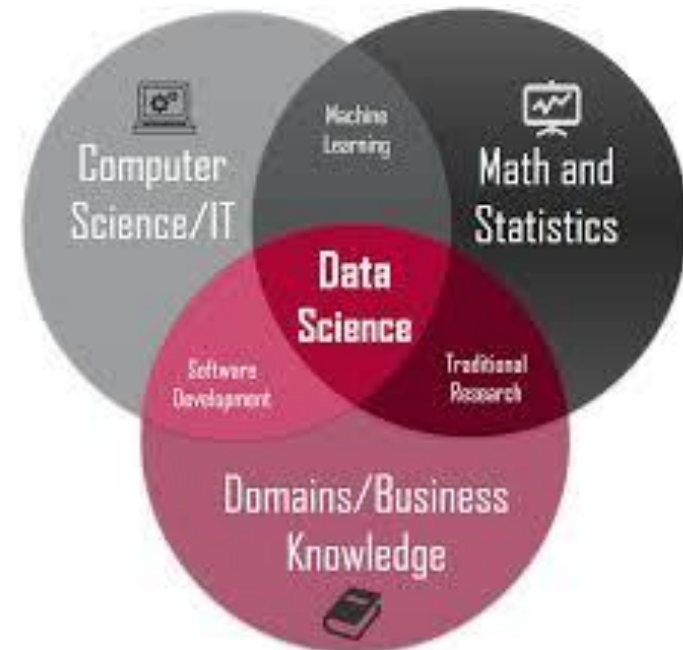
La ciencia de los datos

<https://www.neoland.es/blog/que-es-data-science>

¿Qué es Data Science?

Data Science es la ciencia centrada en el estudio de los datos. Se encarga de extraer información de grandes cantidades de datos.

Data Science combina la estadística, las matemáticas y la informática para interpretar datos. El objetivo es tomar decisiones.



- Estos datos se obtienen a través de diferentes canales. Los teléfonos móviles, las redes sociales, los e-commerce o las encuestas son solo algunas de las fuentes utilizadas.
- **Nuestros gustos, rutinas o movimientos generan datos de gran valor para las empresas que quieren conocer a sus clientes al detalle.**



Sin embargo, la interpretación de los datos no estructurados no aporta valor a las compañías. **De ahí surge la necesidad de contar con científicos de datos en sus equipos.** Gracias al Data Science las empresas pueden anticiparse a la hora de tomar decisiones.



¿De dónde proviene el término Data Science?

- El término 'Data Science' ha estado presente durante las últimas tres décadas.
- No fue hasta la década de los 70 cuando el término se comenzó a usar para definir los métodos de procesamiento de datos.
- Finalmente, 2001 fue el año en el que la ciencia de datos se introdujo como una disciplina independiente.

¿Data Science es lo mismo que Big Data ?

- Cuando hablamos de **Big Data** nos referimos a **conjuntos de datos de gran volumen**. Esto dificulta su almacenaje, gestión, procesamiento y análisis mediante tecnologías y herramientas convencionales.
- En definitiva, el **Big Data se encarga de resolver los problemas de gestión y almacenamiento de datos**. Esto permite dibujar patrones y obtener una visión más completa de los clientes.
- Por otro lado, **Data Science se centra en las herramientas que transforman los datos en información de valor**.

Conceptos clave en Data Science

Data Mining (Minería de Datos)

- Data Mining se define como un **proceso utilizado para recolección y almacenamiento de datos útiles.**
- Para ello es necesario **analizar patrones de datos** en grandes lotes usando aplicaciones que generan informes de resumen.
- Gracias a este proceso, las empresas pueden obtener más información sobre sus clientes y desarrollar estrategias más efectivas :

Decisiones basadas en información.



Para segmentar los datos y evaluarlos, el Data Mining utiliza **algoritmos matemáticos**.

Data mining techniques

Classification

Clustering

Regression

Outer

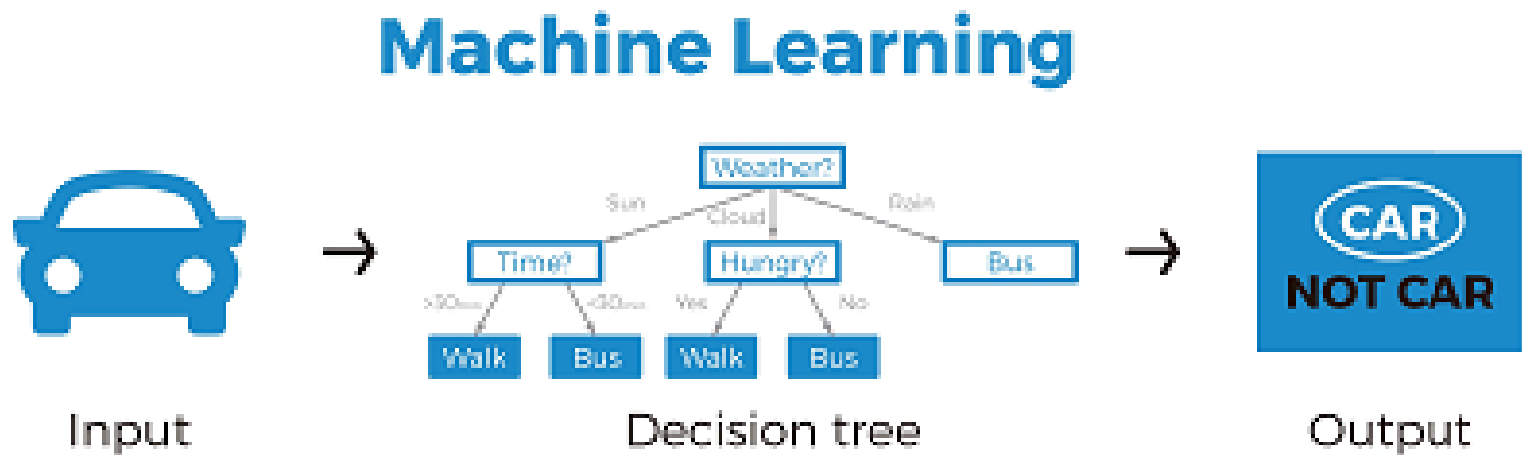
Sequential
Patterns

Prediction

Association
Rules

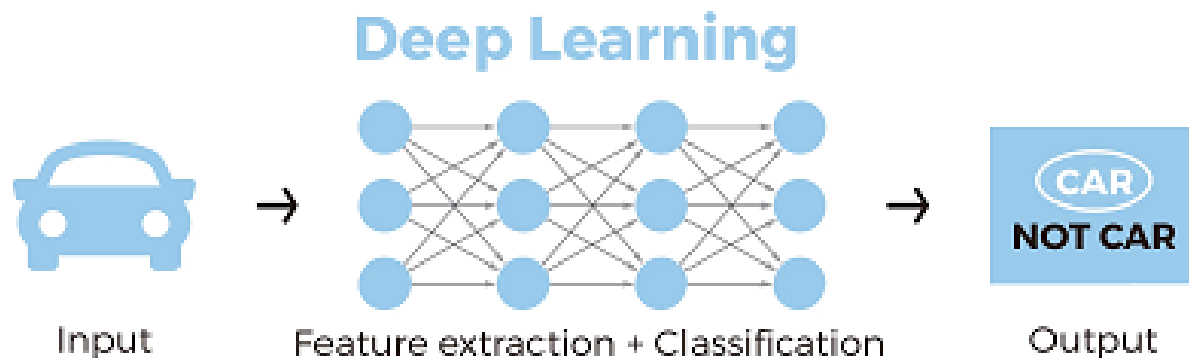
Machine Learning

- Fundamentalmente, Machine Learning se encarga de **educar a la tecnología para que corrija errores** por si sola. Se basa en la **predicción y clasificación de datos** para obtener información útil aplicable a diferentes áreas



Deep Learning

- El objetivo del Deep Learning es **resolver problemas a través de redes neuronales** que imitan el comportamiento del cerebro.
- Estas redes de neuronas artificiales se estructuran en capas.
- En la primera capa es donde se capta la información. Estos datos pasan a la siguiente capa, encargada de realizar cálculos. Y por último, la información recopilada se proyecta en la última de las capas.
- Algunas de las aplicaciones más usadas en Deep Learning son el **procesamiento de textos y el reconocimiento de imagen, objetos o voz.**



Machine Learning



Input



Decision tree

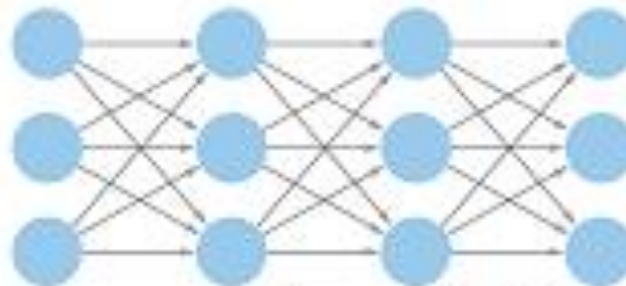


Output

Deep Learning



Input



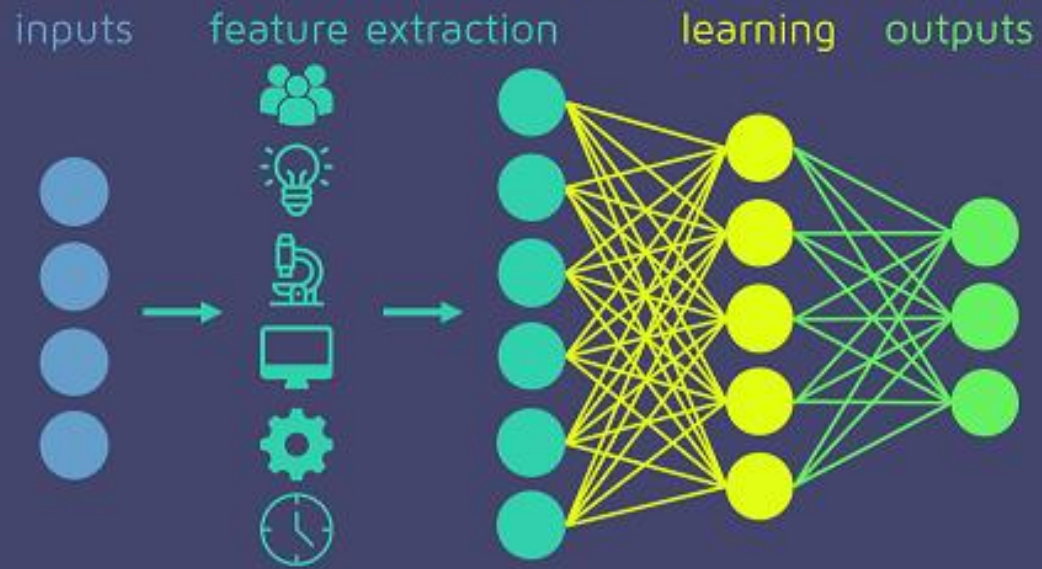
Feature extraction + Classification



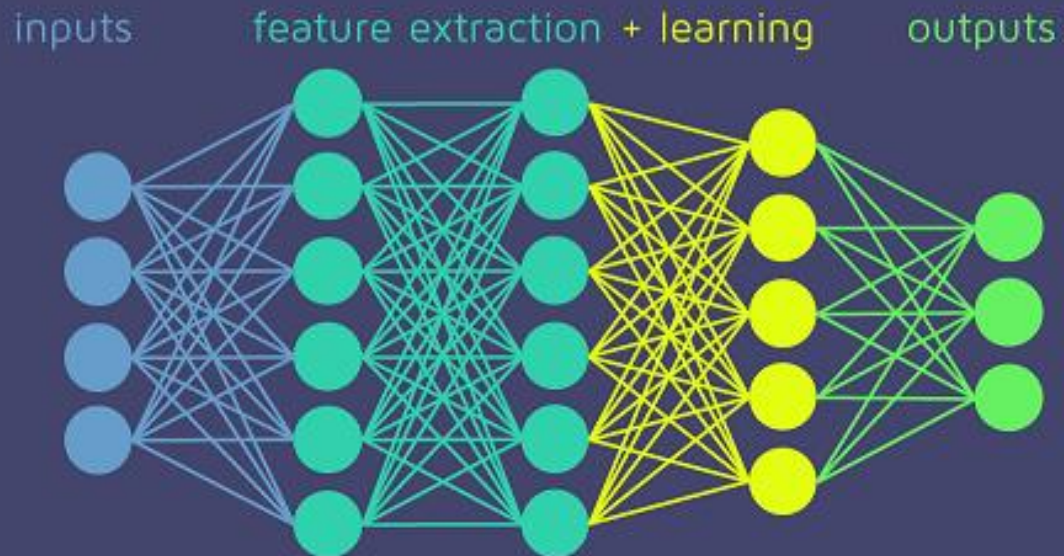
Output

Potenciales de acción -> activación

MACHINE LEARNING

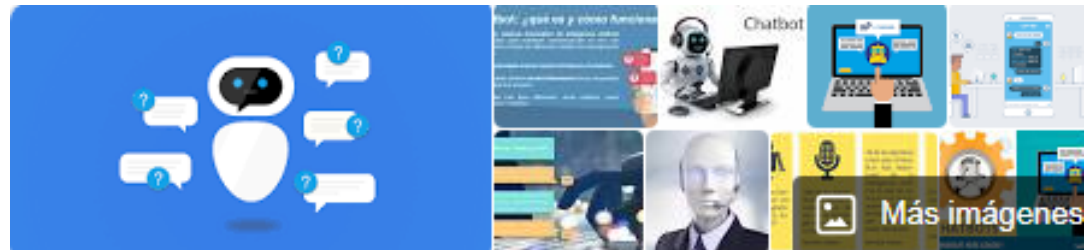


DEEP LEARNING



Inteligencia artificial (IA)

- Se basa en **algoritmos utilizados para la creación de máquinas que imitan el comportamiento humano**. Hoy en día, la inteligencia emocional se aplica en el **reconocimiento facial** o en **la creación de chatbots**, entre otros.



Bot conversacional



Un bot de charla o bot conversacional es un programa que simula mantener una conversación con una persona al proveer respuestas automáticas a entradas hechas por el usuario. Habitualmente, la conversación se establece mediante texto, aunque también hay modelos que disponen de una interfaz de usuario multimedia.

Puestos de trabajo relacionados

Data Scientist

Especializado en la extracción de conocimiento a través de los datos. Entre sus funciones está la **construcción de algoritmos que faciliten la extracción y organización de la información**. Habitualmente, el Data Scientist tiene conocimientos matemáticos, estadísticos y de programación.

Deep Learning

Hace que los datos sean accesibles para su manipulación por parte de los Data Scientist. Entre sus funciones están la utilización de herramientas y **procesos para el desarrollo de aplicaciones de Big Data**.

Puestos de trabajo relacionados

Data Analyst

Es la persona encargada de interpretar los datos extraídos a través de diferentes técnicas.

En definitiva, **analiza y presenta resultados** que deben traducirse en conclusiones.

Chief Data Officer (CDO)

Dentro de una compañía es el **máximo responsable de la gestión de datos**.

Es la persona encargada de liderar al resto de perfiles descritos anteriormente.

Perfil del Data Scientist

Habilidades :

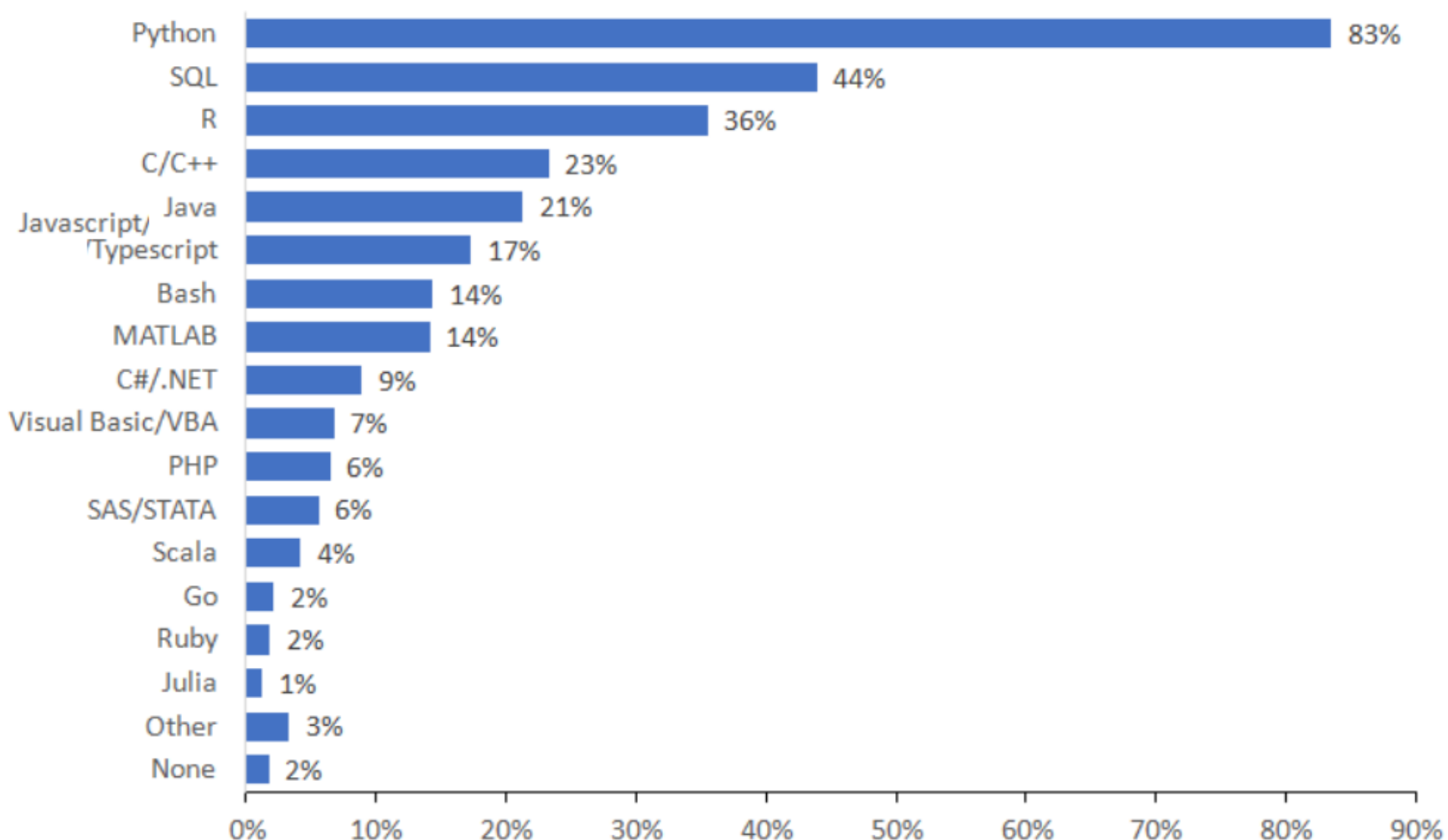
- Recopilar, procesar y extraer valor de las diversas y extensas bases de datos.
- Imaginación para comprender, visualizar y comunicar sus conclusiones a los no científicos de datos.
- Capacidad para crear soluciones basadas en datos que aumentan los beneficios, reducen los costos.
- Los científicos de datos trabajan en todas las industrias y hacen frente a los grandes proyectos de datos en todos los niveles.

Tareas Data Scientist

El proceso que sigue un científico de datos para responder cuestiones que se le plantean se puede resumir en estos pasos:

1. Extraer datos, independientemente de la fuente y de su volumen.
2. Limpiar los datos, para eliminar lo que pueda sesgar los resultados.
3. Procesar los datos usando métodos estadísticos como inferencia estadística, modelos de regresión, pruebas de hipótesis, etc.
4. Diseñar experimentos adicionales en caso de ser necesario.
5. Crear visualizaciones gráficas de los datos relevantes de la investigación.

What programming language do you use on a regular basis?



¿Cuáles son las etapas de un proceso de análisis de datos?

1. **Definir El Problema.** Todo nace de un problema, de una necesidad real. En esta etapa se define el foco del estudio.
2. **La Recolección.** Es el diseño de un método para obtener DATOS : una encuesta, pruebas en laboratorios, datos de marketing, redes sociales etc...
3. **La Limpieza** Homogeneizar los datos en cuanto a formato, deshacer observaciones que no te interesan y almacenar las más útiles.
5. **La Exploración.** Utilizar la estadística descriptiva para traducir los DATOS a gráficos y características que se puedan interpretar.

6. El Análisis. Con la ayuda de la estadística inferencial encontrar conclusiones de un conjunto grande (población) con la información de una pequeña parte de este conjunto (muestra).

>> El contraste de hipótesis es la herramienta más usada.

>> Otras técnicas como: la predicción, la clasificación o

los métodos de causa-efecto.

7. La Conclusión Interpretar los resultados del análisis y listar las conclusiones.

<https://conceptosclaros.com/como-analizar-datos/>

6 problemas más comunes

Distinguir si un grupo es diferente a otro. Por ejemplo “ver si la presión arterial es más alta en un grupo de pacientes con obesidad y los que no” Por un lado tienes la variable numérica (presión arterial) y por el otro grupos de pacientes (con o sin obesidad)

Distinguir si las proporciones son diferentes de un grupo al otro. Por ejemplo: “*¿La proporción de accidentes de moto es diferente por rangos de edad 15-20, 20-25, 25-30 y 30-35?*» Tenemos 4 proporciones que queremos comparar.

Ver si los grupos tienen relación en la tabla de contingencias. Por ejemplo “*¿Hay relación en cuanto a la precisión (buena, regular, mala) de un aparato de medida analógico y uno óptico?*” En el fondo, es relacionar variables cualitativas.

Analizar si hay relación entre variables numéricas. La asociación entre medidas. Por ejemplo “*Tienes datos de 200 familias sobre las ganancias y los gastos. ¿Es verdad que se cumple que cuánto más tienes más gastas?*”

Calcular un modelo matemático que permita predecir una variable en función de otras. Por ejemplo “*Un modelo que te permita calcular la frecuencia cardíaca máxima en función de datos como la edad, la presión baja y alta, el sexo, la altura y el peso*”

Técnicas avanzadas de reconocimiento de patrones o machine learning. Modelos y algoritmos predictivos para poder aportar respuestas de tus datos y solucionar problemas reales con aplicaciones personalizadas para tu caso.