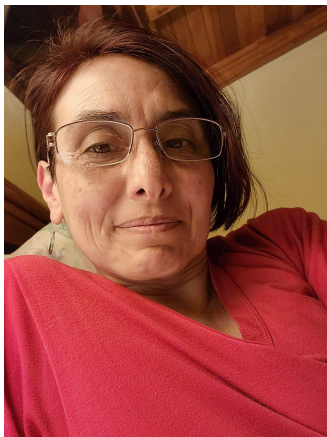


Análisis y Visualización de Datos

Diplomatura CDAAyA 2024

Profes



Georgina Flesia

Profe en FaMAF - UNC
Lic. en Matemática y Dra. en Matemática.
Trabajo en Procesamiento de Señales y
Sistemas

georgina.flesia@unc.edu.ar



Karim Nemer

Ing en Sistemas de Información UTN-FRC,
Dra en Ingeniería, mención Electrónica UTN-FRC
Trabajo en Inteligencia Artificial aplicado al
procesamiento de imágenes satelitales

karim.nemer@unc.edu.ar

APLICACIONES

Recomendadores en base a
comportamientos
Netflix - Spotify - Amazon



Procesamiento de lenguaje
natural



Detección de enfermedades

Cáncer basada en imágenes -
Predicción de epidemias

Chatbots
Asesores
de
compraS



Clasificación de
emails Spam - Correo
no deseado



Reconocimiento Facial,
Voz, Huellas



APLICACIONES

**Chatbot-ANÁLISIS DE
TURNOS MÉDICOS**

**Personalización
de Servicios
bancarios**

**Como hacer un clasificador
de pliegos?**

**Cambio climático- análisis de
datos de emisión de gases**

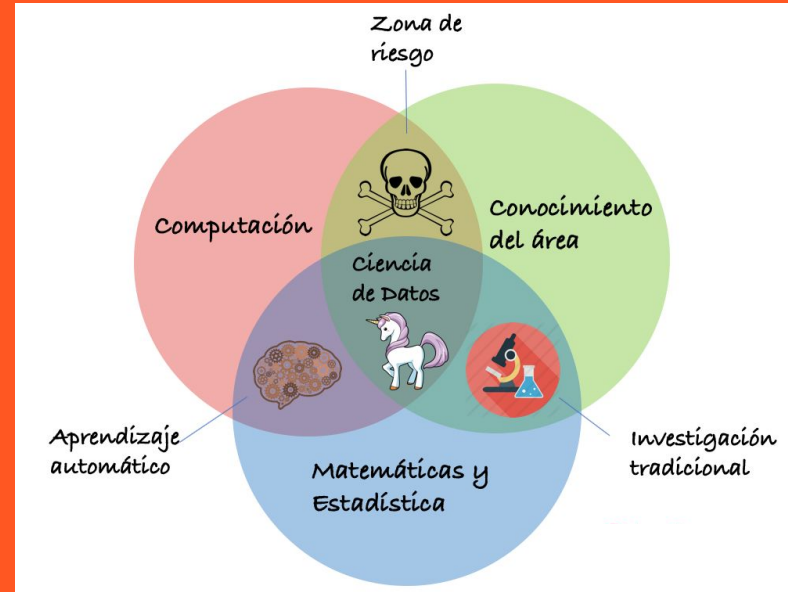
**Descifrando el
futuro económico:
inflación,
recomendación de
activos financieros**

**Reconocimiento de
patologías en
imágenes**

**Detección de sentimientos
en foros
Ansiedad, depresión,
agresión**

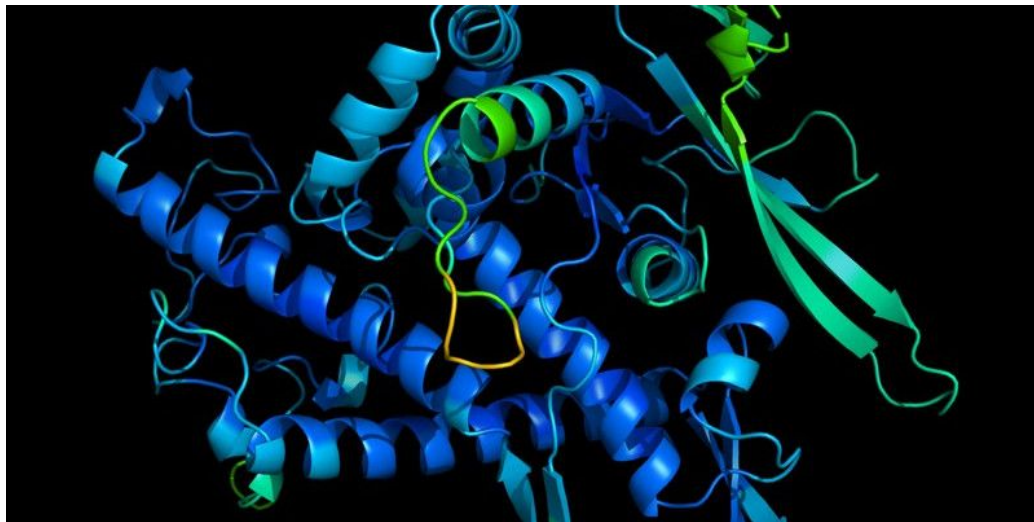
¿Qué es la Ciencia de Datos?

Conjunto de disciplinas que utilizan métodos científicos para obtener conocimientos a partir de los datos dándoles un mayor valor.



AlphaFold

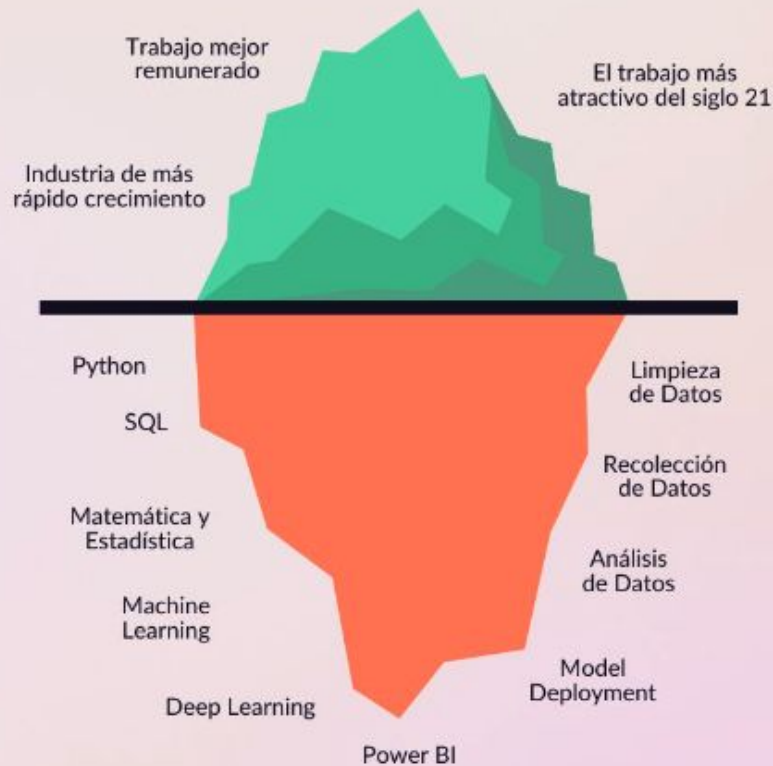
AlphaFold, predicción de estructuras 3D de proteínas mediante IA



<https://www.youtube.com/watch?v=58XzFvZ8xVU>

DATA SCIENCE

LO QUE VEMOS vs. LO QUE REALMENTE ES



Como se trabaja

Data Analyst

Parte de preguntas
concretas

Busca explicar los datos
para tomar decisiones

Guiado por la intuición
del analista

Detecta patrones
superficiales

Data Science

Machine Learning

Como se trabaja

Data Analyst

Parte de preguntas concretas

Busca explicar los datos para tomar decisiones

Guiado por la intuición del analista

Detecta patrones superficiales

Data Science

Parte de una situación problemática

Busca un producto de datos

Guiado por interpretación de resultados

Hace emerger patrones profundos

Machine Learning

Como se trabaja

Data Analyst

Parte de preguntas concretas

Busca explicar los datos para tomar decisiones

Guiado por la intuición del analista

Detecta patrones superficiales

Data Science

Parte de una situación problemática

Busca un producto de datos

Guiado por interpretación de resultados

Hace emerger patrones profundos

Machine Learning

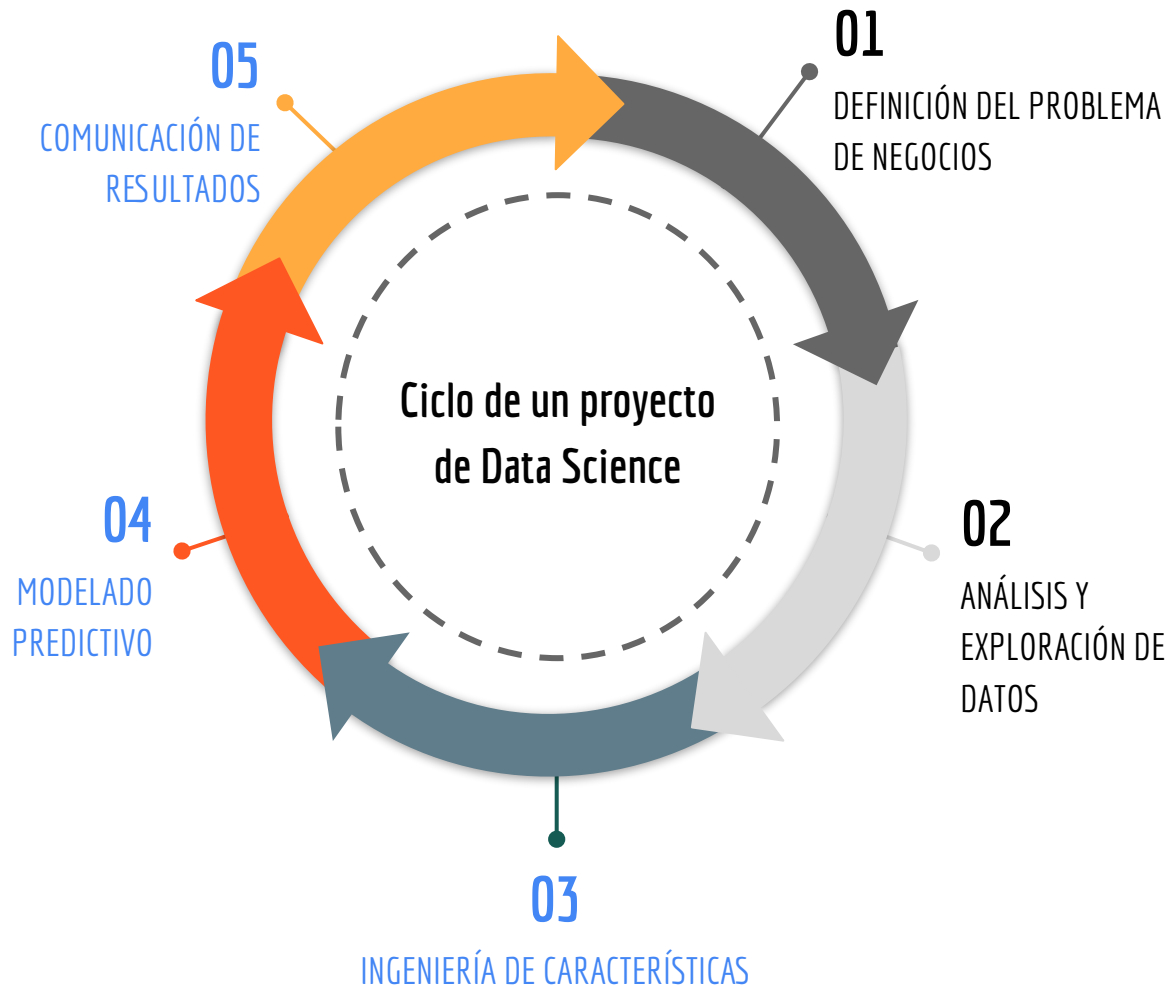
Parte de una tarea y un conjunto de datos

Busca optimizar una métrica de desempeño

Guiado por la teoría de los modelos

Detecta patrones profundos

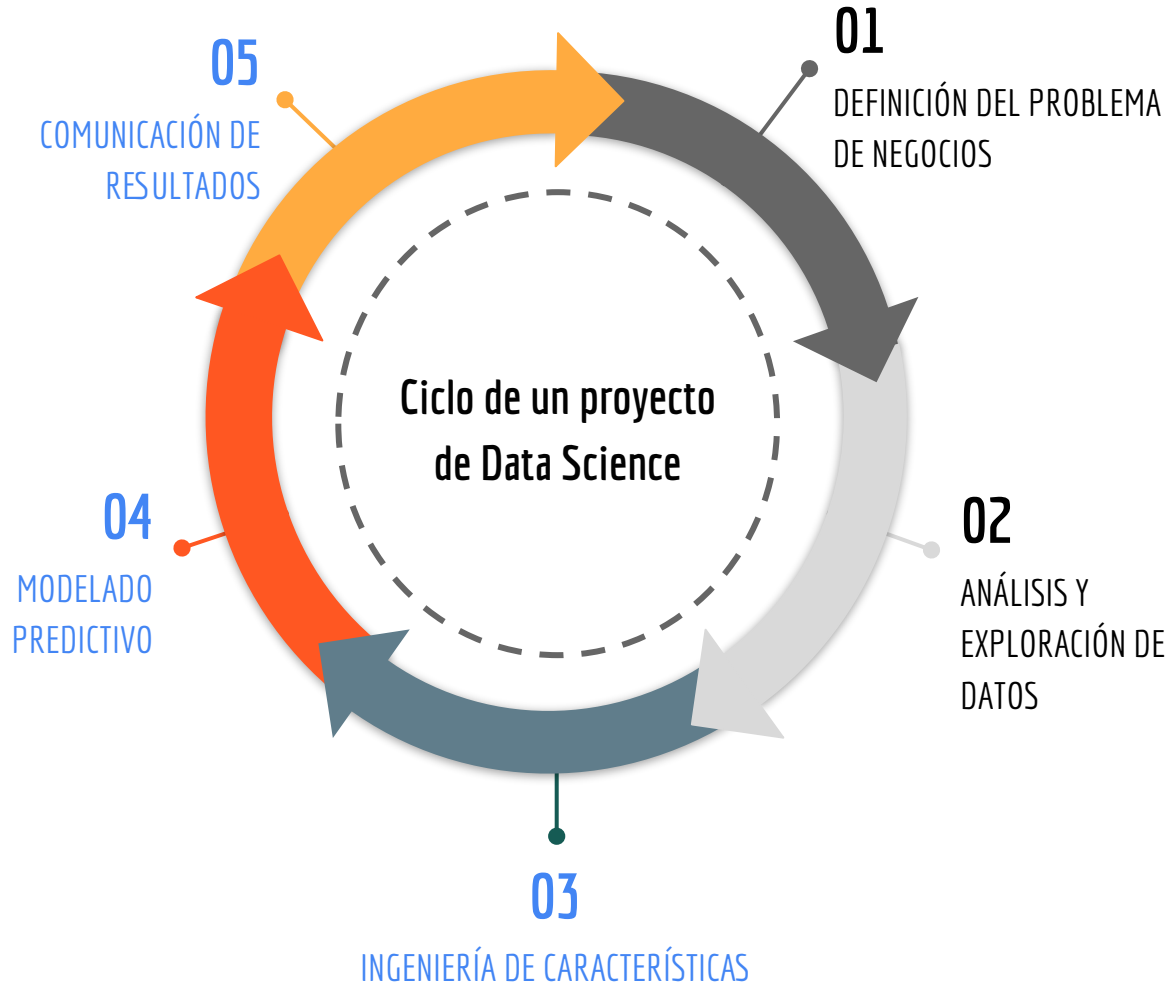
El ciclo sin fin



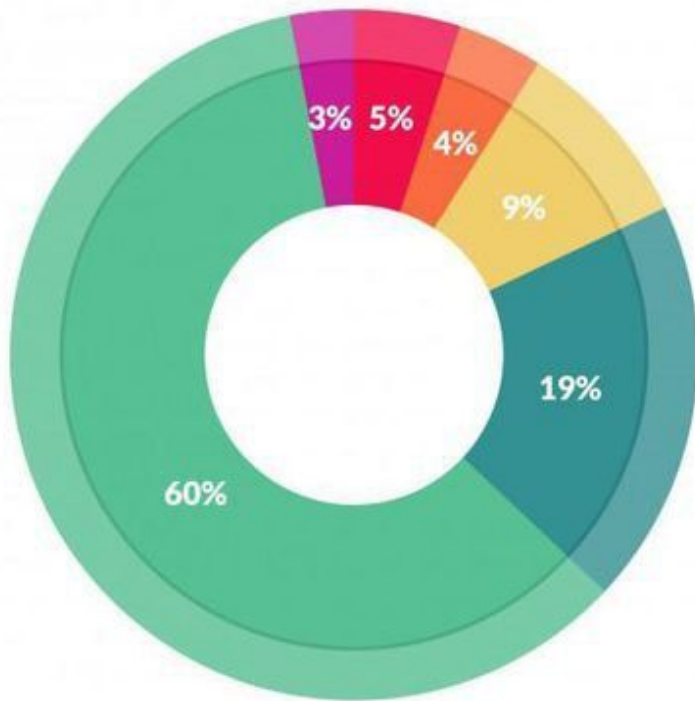
El ciclo sin fin

Durante esta materia, veremos conceptos involucrados en:

1. Herramientas estadísticas y visualizaciones para la **etapa 02.**
2. Herramientas estadísticas necesarias para interpretar los resultados de la **etapa 04.**
3. Visualización y comunicación efectiva para la **etapa 05.**



REALIDAD



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Sesgos en los datos

¿Qué observa en esta imagen?



Sesgos en los datos

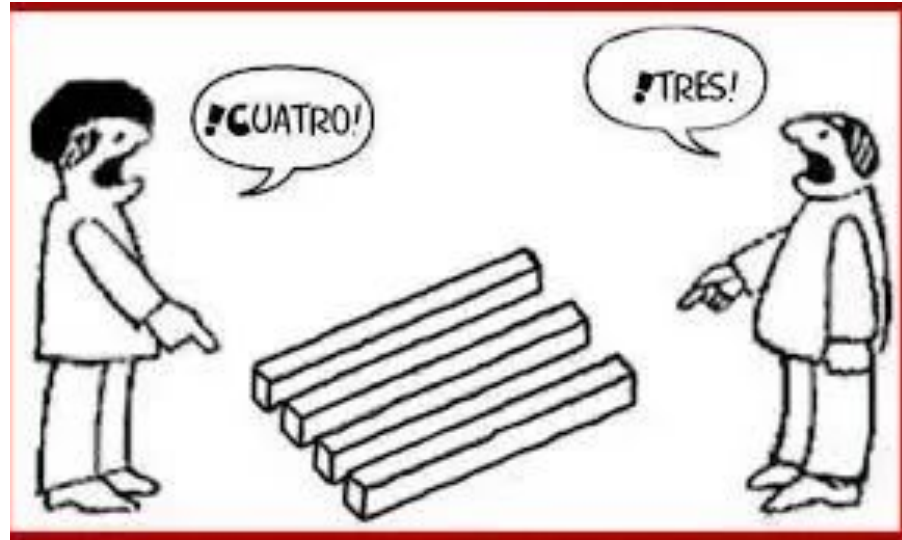
¿Qué observa en esta imagen?

- Bananas
- Etiquetas
- Bananas en estantes






Sesgos en los datos

El sesgo no proviene de algoritmos de IA, proviene de personas.



De que se está hablando...

ChatGPT		
 Examples	 Capabilities	 Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

¿Qué puedes hacer con ChatGPT?

Esta herramienta es totalmente versátil, ya que puedes pedirle que te explique cualquier **concepto**, que escriba **artículos o resúmenes** de cualquier suceso histórico, o que invente **poemas y chistes**. Además, podrás pedir consejo sobre qué teléfono comprar, así como triviales, **traducciones**, definiciones, explicaciones, y todo lo que se te ocurra.

Conversatorio sobre Inteligencia Artificial y ChatGPT- Laura Alonso Alemany

https://www.youtube.com/watch?v=0R6Ysg4t_Sk&t=13s

Habilidades a aprender durante esta materia

1 *Seleccionar y aplicar herramientas estadísticas **adecuadas***

2 *Diseñar procesos de análisis de datos **sistemáticos***

3

4

5

Habilidades a aprender durante esta materia

- 1** *Seleccionar y aplicar herramientas estadísticas **adecuadas***
- 2** *Diseñar procesos de análisis de datos **sistemáticos***
- 3** *Obtener resultados a partir de un conjunto de datos y **contextualizarlos***
- 4** *Explicar resultados y conclusiones de forma **correcta** y **efectiva***
- 5**

Habilidades a aprender durante esta materia

- 1** *Seleccionar y aplicar herramientas estadísticas **adecuadas***
- 2** *Diseñar procesos de análisis de datos **sistemáticos***
- 3** *Obtener resultados a partir de un conjunto de datos y **contextualizarlos***
- 4** *Explicar resultados y conclusiones de forma **correcta y efectiva***
- 5** *Implementar pipelines de análisis de datos en Python*

¿Qué pasa si usamos Machine Learning sin saber análisis de datos?

- No sabemos **qué modelar**, a menos que alguien más nos diga cómo.

¿Qué pasa si usamos Machine Learning sin saber análisis de datos?

- No sabemos qué modelar, a menos que alguien más nos diga cómo.
- **No podemos interpretar correctamente el impacto de sus resultados**
 - **Impacto a largo plazo, provocados por sesgos, unfairness, filtrado de información**
 - **Impacto en métricas de negocios, por ejemplo evaluado a través de test A/B**

¿Qué pasa si usamos Machine Learning sin saber análisis de datos?

- No sabemos **qué modelar**, a menos que alguien más nos diga cómo.
- No podemos interpretar correctamente el **impacto** de sus resultados
 - Impacto a largo plazo, provocados por sesgos, unfairness, filtrado de información
 - Impacto en métricas de negocios, por ejemplo evaluado a través de test A/B
- **Perdemos mucho tiempo en desarrollar modelos que no responden la pregunta correcta, y por lo tanto son menos accionables**

¿Qué sucede si hacemos Data Science sin entender Machine Learning?

- **Estamos limitados a análisis simples. O usamos modelos sin saber cómo funcionan...**

¿Qué sucede si hacemos Data Science sin entender Machine Learning?

- Estamos limitados a análisis simples. O usamos modelos sin saber cómo funcionan...
- **No entendemos qué tipo de restricciones imponen los modelos que elegimos, por ejemplo, modelos lineales, hiperplanos, etc.**

¿Qué sucede si hacemos Data Science sin entender Machine Learning?

- Estamos limitados a análisis simples. O usamos modelos sin saber cómo funcionan...
- No entendemos qué tipo de **restricciones imponen los modelos** que elegimos, por ejemplo, modelos lineales, hiperplanos, etc.
- **Aplicamos modelos a conjuntos de datos para los cuales no son adecuados, por ejemplo, redes neuronales sin normalizar las columnas.**

¿Qué sucede si hacemos Data Science sin entender Machine Learning?

- Estamos limitados a análisis simples. O usamos modelos sin saber cómo funcionan...
- No entendemos qué tipo de **restricciones imponen los modelos** que elegimos, por ejemplo, modelos lineales, hiperplanos, etc.
- Aplicamos modelos a conjuntos de datos para los cuales no son adecuados, por ejemplo, redes neuronales sin normalizar las columnas.
- **Perdemos mucho tiempo en optimizar los modelos, porque no sabemos cómo hacerlo.**

Material y Herramientas de trabajo

- Google colab → Para leer notebooks de python (00 Inicios en Python.ipynb)
- Aula Virtual → Nuestro depósito de datos y herramientas, restringido sólo a los alumnos de la Diplomatura



Inciando...

00 Inicios en Python.ipynb

