

Ciencia de Datos en el mundo empresarial

“Nueva forma de generar valor a la empresa”

OSCAR ANDRES GASPAR ALVAREZ

Economista y Estudiante de maestría Métodos cuantitativos

LINKEDIN:<https://www.linkedin.com/in/oscar-andres-gaspar-alvarez-12a173133/>

YOUTUBE:https://www.youtube.com/channel/UCbyxBA6pOFyak50EpCWZj9g?view_as=subscriber

CEL:3217291122

```

r_mod.use_x = False
ration == "MIRROR_Y"
or_mod.use_x = False
or_mod.use_y = True
or_mod.use_z = False
eration == "MIRROR_Z":
or_mod.use_x = False
r_mod.use_y = False
r_mod.use_z = True
mirror_mod.use_x =
class MirrorModData vs.
action at the end - add
mirror_mod.use_x = 1
_ob.select = 1
fier_ob.select = 1
.context.scene.objects.acti
nt("Selected" + str(modifi
#mirror_ob.select = 0
bpy.context.selected ob
ta.object.name == 1, se
mirror_mod
r_mod = modifier_ob.a
elif_operatio
nor object to mirror
mod.mirror_object
n == "MIRROR_X":
od.use_x = True
' use v = False
```

Objetivos.

- El rol del científico de datos en las compañías.
- Metodologías de proyectos de Analítica en las empresas.
- ¿Qué es la inteligencia artificial?
- ¿Qué es Machine learning?
- Componentes del Machine Learning

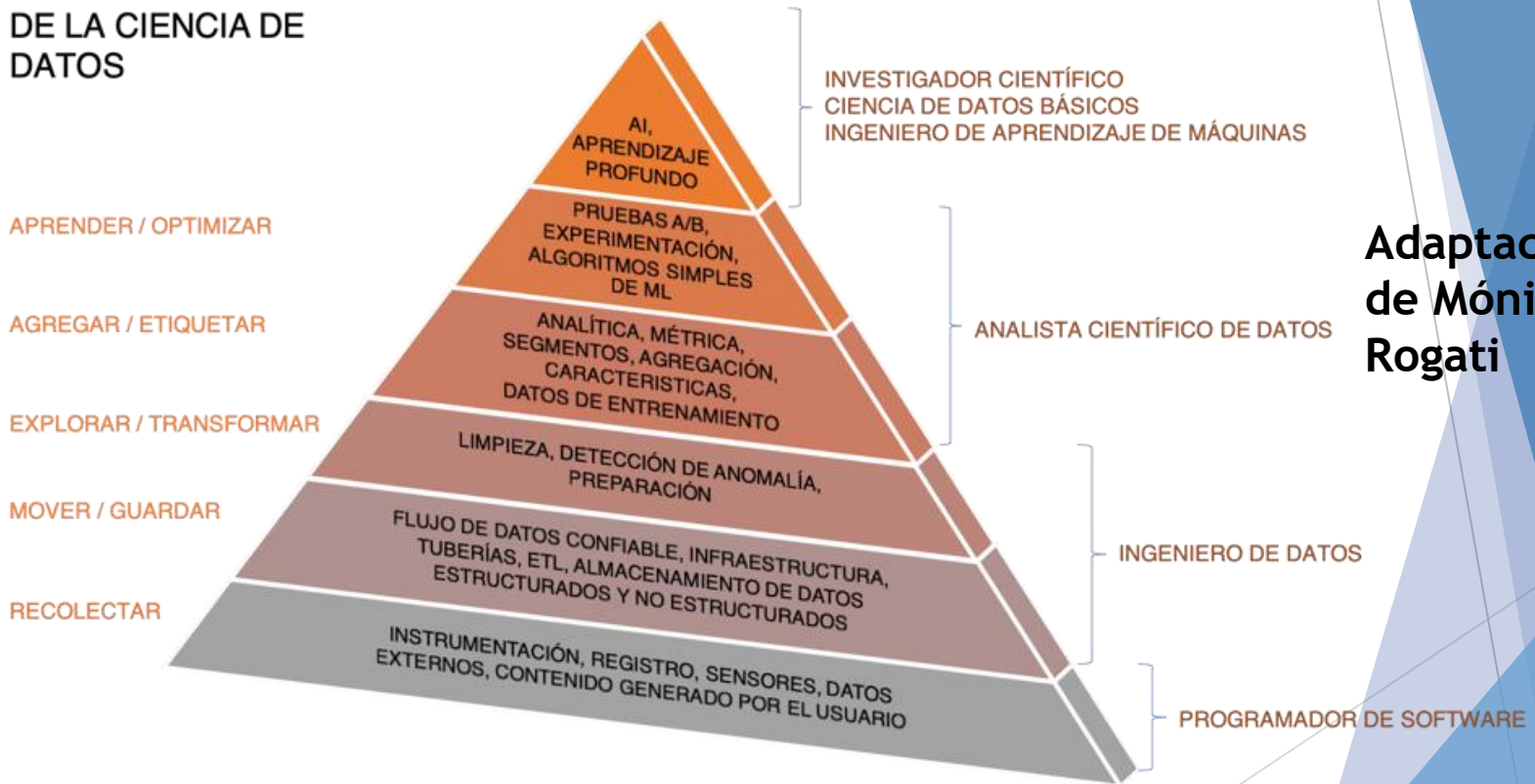
El rol del científico de datos en las compañías

- ▶ Ser un Científico de datos no se trata de cuan avanzados son tus modelos, se trata de cuanto impacto puede tener tu trabajo.
- ▶ Eres un solucionador de Problemas. Un estrategia de la empresa.



Grandes Empresas

LA JERARQUÍA DE NECESIDADES DE LA CIENCIA DE DATOS



Adaptación
de Mónica
Rogati

Empresas Medianas

THE DATA SCIENCE HIERARCHY OF NEEDS

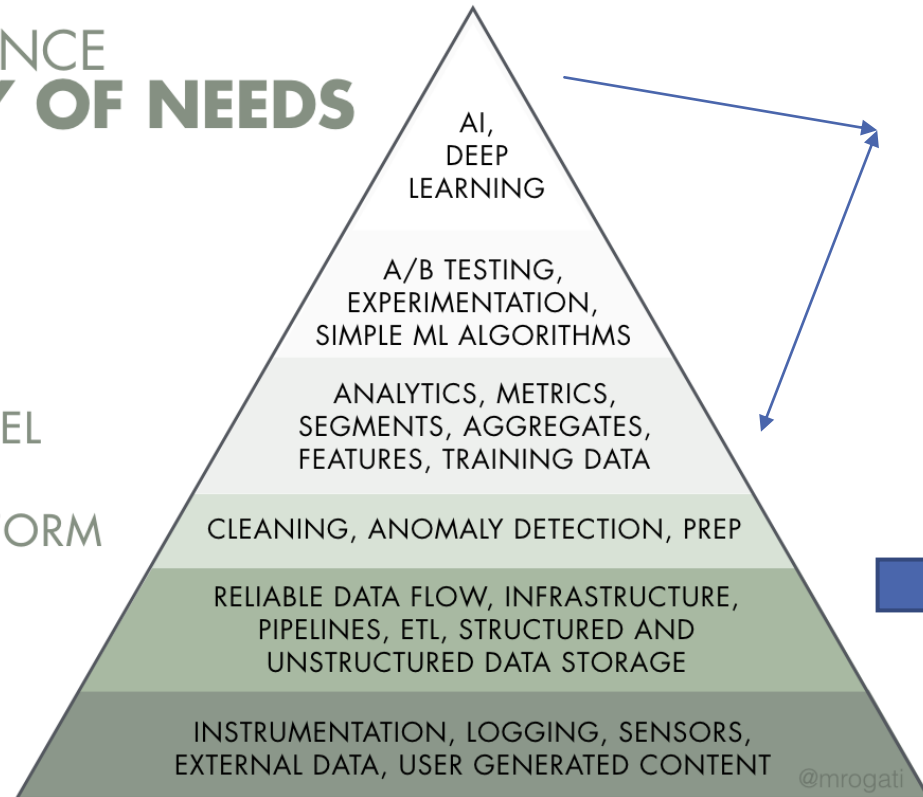
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



DS

DE

SE

Empresas Pequeñas

THE DATA SCIENCE HIERARCHY OF NEEDS

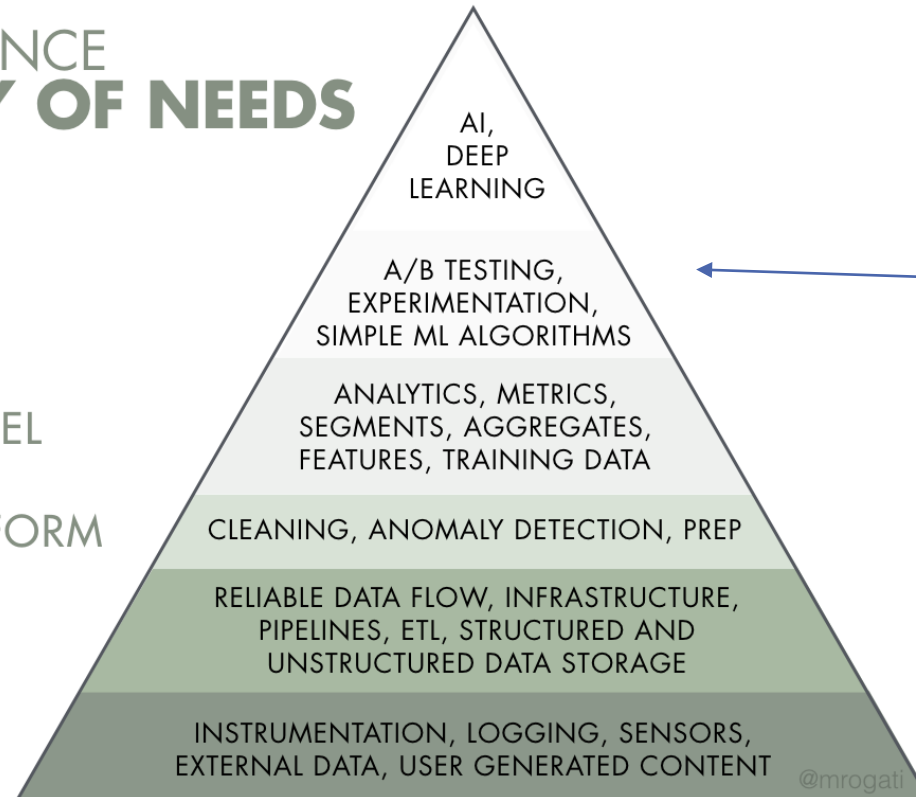
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

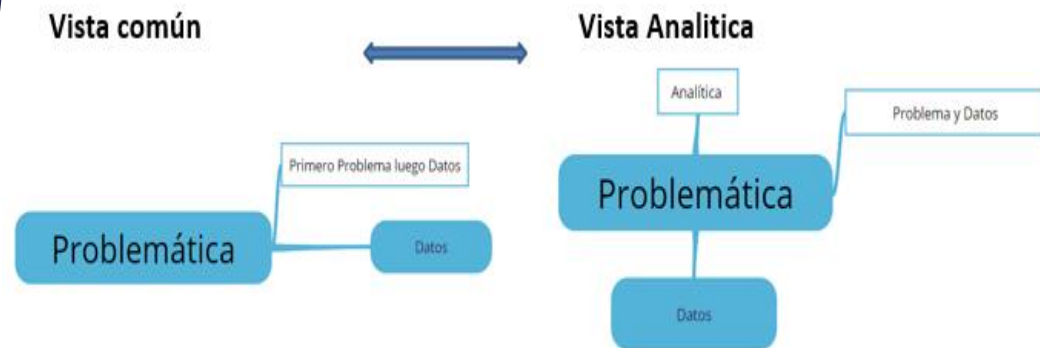


**Data
Scientist.**

Definir problemática.

- El propósito general de todo proyecto es aportar a la solución de una problemática.
- ¿Qué es problemática ?

Aquella situación sobre la cual no se tiene control y no es evidente tomar acciones confiables.



1) Claves para definir problemática

Entender el problema es reconocer la falta de información que impide tomar decisiones acertadas

¿cuáles son las variables que hicieron evidente el problema?

¿Cuáles son las variables que influyen en el problema observado?

Ejemplos de problemática : no alcanzar la meta de ventas, querer segmentar los clientes, caracterizar el mercado, señalar oportunidades de crecimiento.

Preguntas como :

Por qué, cómo, cuándo, dónde, cuánto, cuál, entre otras.

Metodologías de Proyectos de Analítica

- ▶ Establecer una forma probada para llevar un problema a una solución de analítica
- ▶ Lograr eficiencia.
- ▶ Tomar decisiones en el momento adecuado



Metodologías

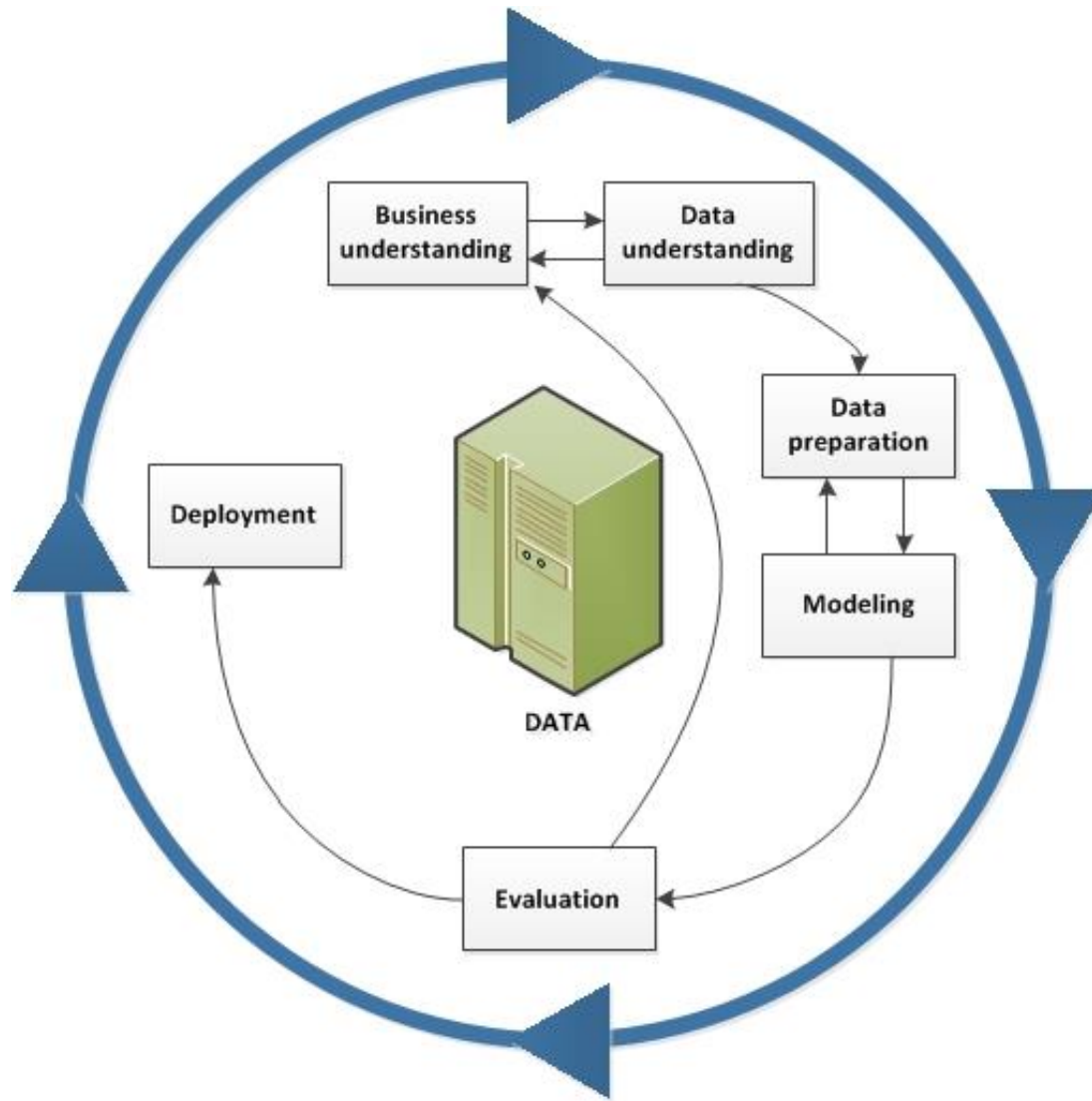
► CRISP-DM

La metodología CRISP-DM (Cross Industry Standard Process for Data Mining) es un modelo estándar abierto propuesto en 1999 por IBM para proyectos relacionados con minería de datos.

► ASUM-DM

se basa en la metodología CRISP-DM y hace énfasis en las nuevas prácticas en la ciencia de datos como el uso de volúmenes de datos muy grandes, la incorporación de análisis de texto, en el modelado predictivo y la automatización de algunos procesos

Metodología CRISP-DM



Business Understanding.

- 1) Definición de necesidades del cliente.
- 2) Conocimiento de los datos en la definición de un problema de minería de datos
- 3) Plan preliminar diseñado para alcanzar los objetivos



Data Understanding



Comienza con la colección de datos inicial y continúa con las actividades que permiten familiarizarse con los datos.



Identificar los problemas de calidad



Descubrir conocimiento preliminar sobre los datos.



Descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta

Data Preparation

- ▶ Construir el conjunto final de datos (los datos que se utilizarán en las herramientas de modelado)
- ▶ **Análisis de los datos y selección de características.**
- ▶ Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación, datos Atípicos, Datos ausentes y la limpieza de datos para las herramientas que modelan.

Modeling

- ▶ En esta fase, se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema (cuantas más mejor), y se calibran sus parámetros a valores óptimos.
- ▶ Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos.
- ▶ Algunas técnicas tienen requerimientos específicos sobre la forma de los datos. Por lo tanto, casi siempre en cualquier proyecto se acaba volviendo a la fase de preparación de datos.

Evaluation

En esta etapa en el proyecto, se han construido uno o comparado el modelo obtenido con los objetivos de negocio.

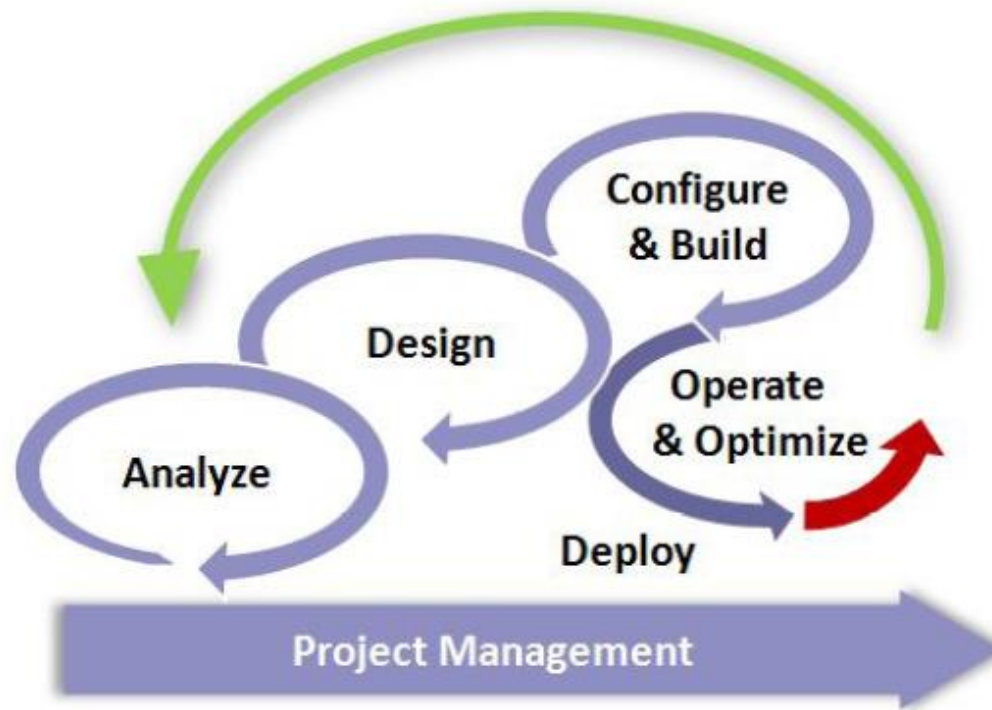
Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no haya sido considerada suficientemente.

Al final de esta fase, se debería obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos.

Deployment

- ▶ La creación del modelo no es el final del proyecto.
- ▶ El conocimiento obtenido tendrá que organizarse y presentarse a la compañía o cliente.
- ▶ Dependiendo de los requisitos, la fase de desarrollo puede ser :
- ▶ Tan simple como la generación de un informe.
- ▶ Tan compleja como la realización periódica y quizás automatizada de un proceso o aplicación.

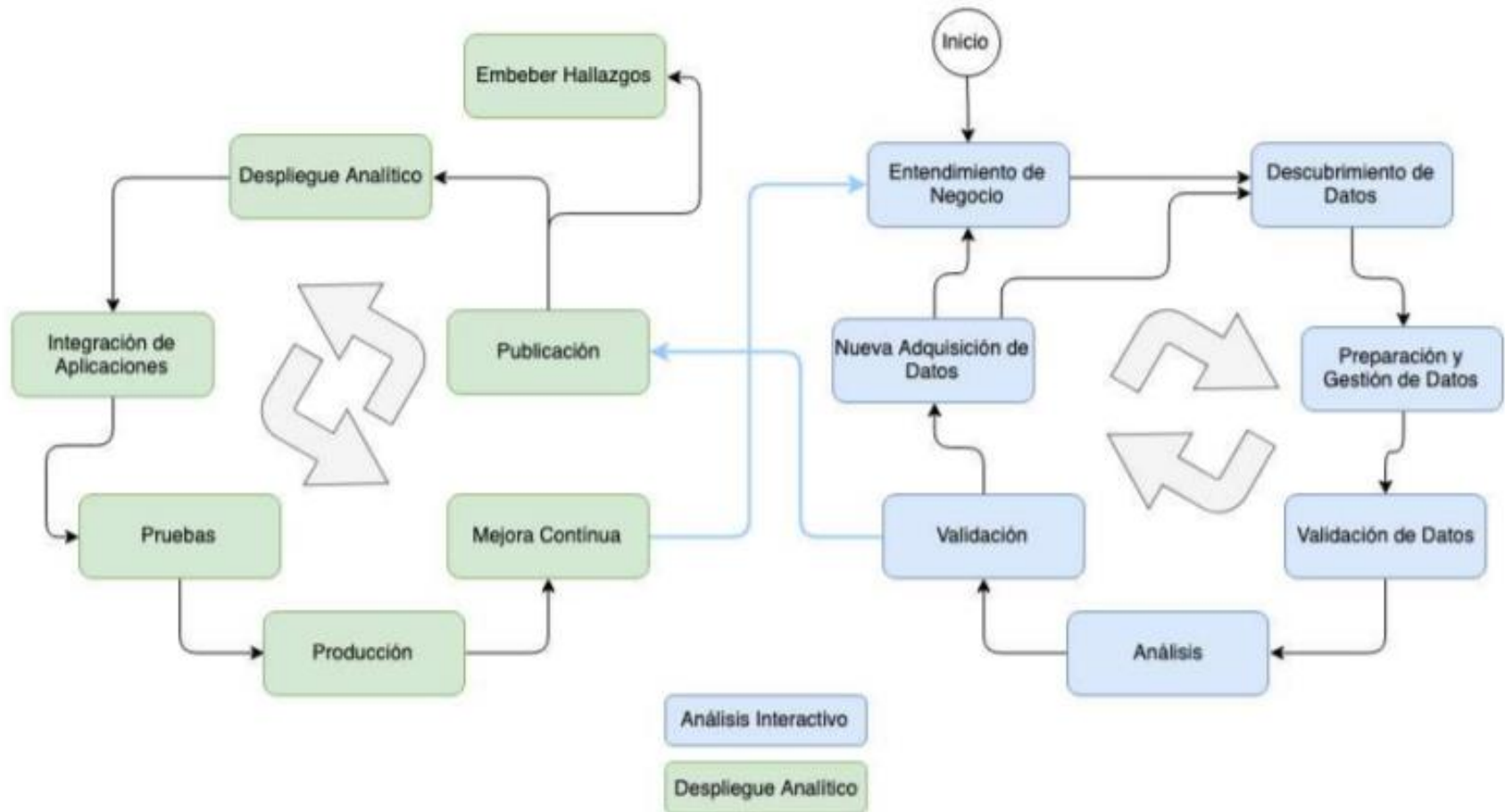
ASUM-DM



Metodología ASUM-DM para analítica en IDECA2019 (Infraestructura de Datos Espaciales de Bogotá)

Ciclo de Publicación

Ciclo de Elaboración



¿Qué es la inteligencia Artificial?

- ▶ La inteligencia artificial (IA) nació en los años 50 aproximadamente 50, algunos pioneros del naciente campo de las ciencias de la computación. Estos personajes comenzaron a preguntarse acerca de poder lograr que los computadores “pensaran”.
- ▶ Alan Turing, matemático británico, por allá en el año 1950 con un trabajo influyente, abre una nueva disciplina , las ciencias de la información.



**Turing propuso la pregunta
¿puede pensar una máquina?**



¿Puede pensar una máquina?

- ▶ El enfoque de Turing de ver a la inteligencia artificial como una imitación del comportamiento humano
- ▶ Fue tan práctico a lo largo del tiempo, y debido a ello la inteligencia artificial focalizo su objetivo hacia el comportamiento racional típicas de reglas.
- ▶ Estudiar el comportamiento humano se dejo a las ciencias cognitivas, era muy difícil decirle a una maquina venga y aprenda como un humano

IA simbólica

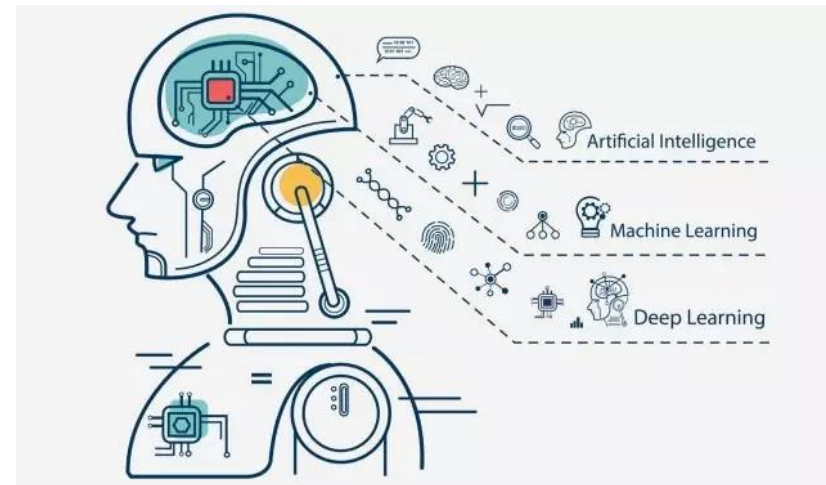
- ▶ Fue el paradigma que dominó el campo de la IA desde 1950 hasta finales de 1980 y alcanzó el pico de su popularidad durante el boom de los Sistemas Expertos en 1998
- ▶ IA con nivel humano podía alcanzarse al hacer que los programadores crearan a mano un conjunto de reglas lo suficientemente grandes para manipular el conocimiento y así generar máquinas inteligente
- ▶ Con el tiempo se volvió intratable el encontrar reglas explícitas para resolver problemas mucho más complejos



Nuevo enfoque : El Machine Learning o Aprendizaje Automático

- ▶ ¿Puede un computador ir más allá de lo que le ordenamos como hacer y aprender por sí mismo como realizar una tarea específica?
- ▶ ¿Podría un computador sorprendernos?
- ▶ ¿podría un computador automáticamente aprender esas reglas directamente de los datos que le pasamos?

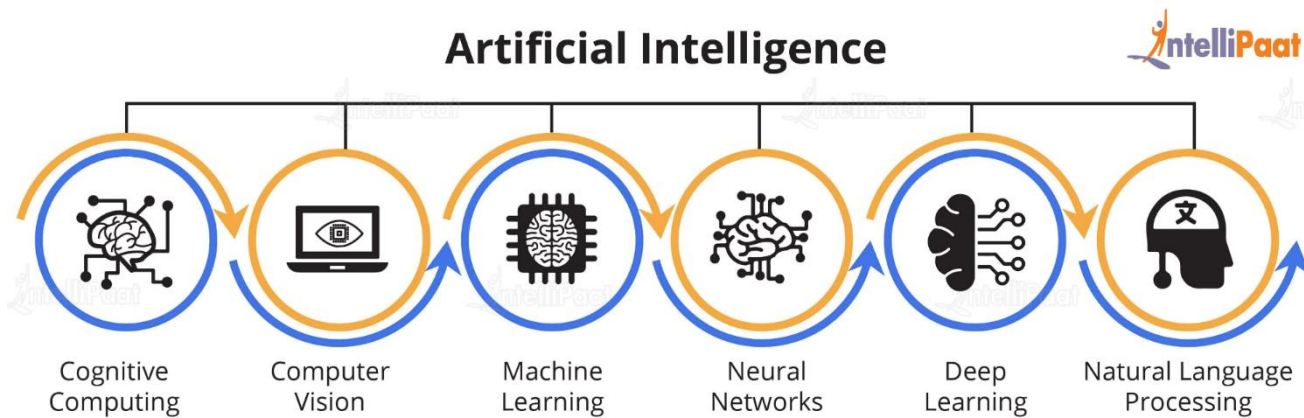
Nuevo enfoque : El Machine Learning o Aprendizaje Automático



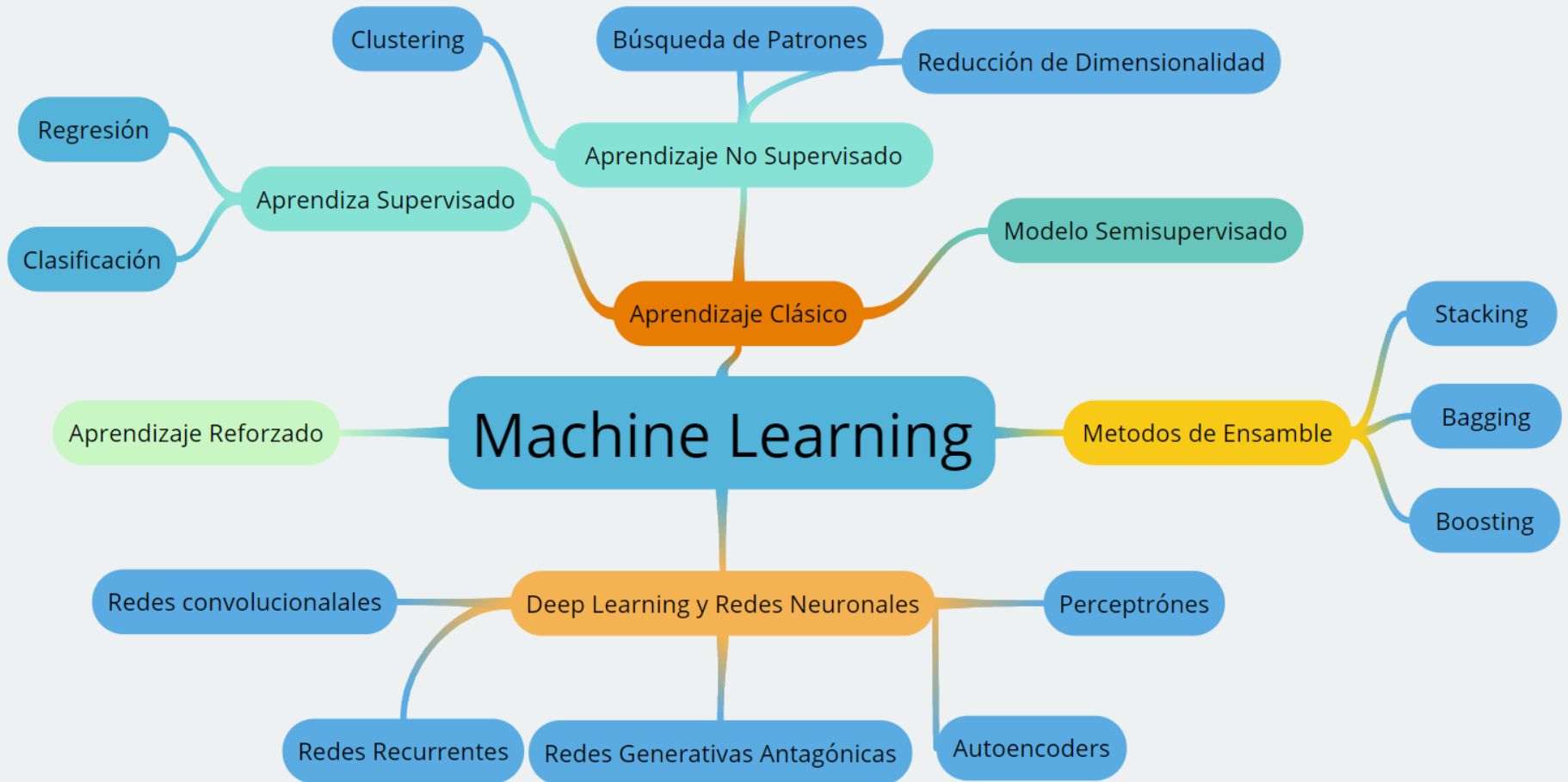
Definición de ML

- ▶ “los humanos pasamos los datos como entrada al igual que las respuestas esperadas de dichos datos con el fin de obtener a la salida las reglas que nos permiten hacer el mapeo efectivo entre entradas y sus correspondientes salidas. Estas reglas pueden ser luego aplicadas a nuevos datos para producir respuestas originales, es decir, generadas automáticamente por las reglas que el sistema “aprendió” y no por reglas explícitamente codificadas por programadores.”
- ▶ Una maquina entonces encuentra la estructura estadística o patrones

Componentes del la Inteligencia Artificial



Componentes ML



Referencias

- ▶ https://www.ideca.gov.co/sites/default/files/MetodologiaAnaliticaDatos_0.pdf
- ▶ https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_crispdm_ddita/elementine/crisp_help/crisp_overview.html
- ▶ <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>
- ▶ <https://code.likeagirl.io/5-componentes-que-la-inteligencia-artificial-debe-poseer-para-tener-%C3%A9xito-480b5deb7dc2>
- ▶ <https://planetachatbot.com/cu%C3%A1les-son-las-principales-diferencias-entre-inteligencia-artificial-y-machine-learning-3ffa6db9e434>
- ▶ <https://cleverdata.io/que-es-machine-learning-big-data/>
- ▶ <https://enpositivo.com/2015/09/inteligencia-artificial-los-10-principales-hitos-del-siglo-xx/>
- ▶ https://books.google.com.co/books?hl=es&lr=&id=dLQ3bDy2tgYC&oi=fnd&pg=PA25&dq=history+of+artificial+intelligence&ots=OBmVay_yzD&sig=7vD4oAS_MOmEPweXzinQujGJcCs#v=onepage&q=history%20of%20artificial%20intelligence&f=false
- ▶ <https://telefonicabusinesssolutionsca.blog/inteligencia-artificial-descubriendo-a-los-pioneros/>