



Laboratorio Nacional de Informática Avanzada A.C.

**CURSO:**

**Taller de Machine Learning  
para el análisis y visualización  
en Power BI**

**Módulo II. Métodos de Machine Learning**

# **Laboratorio de Inteligencia Artificial**

**Tema 01 – Inteligencia Artificial en la predicción**

**Profesor: Saúl Domínguez Isidro, PhD.**

**Contacto: [saul.dominguez@lania.edu.mx](mailto:saul.dominguez@lania.edu.mx)**

# Objetivo

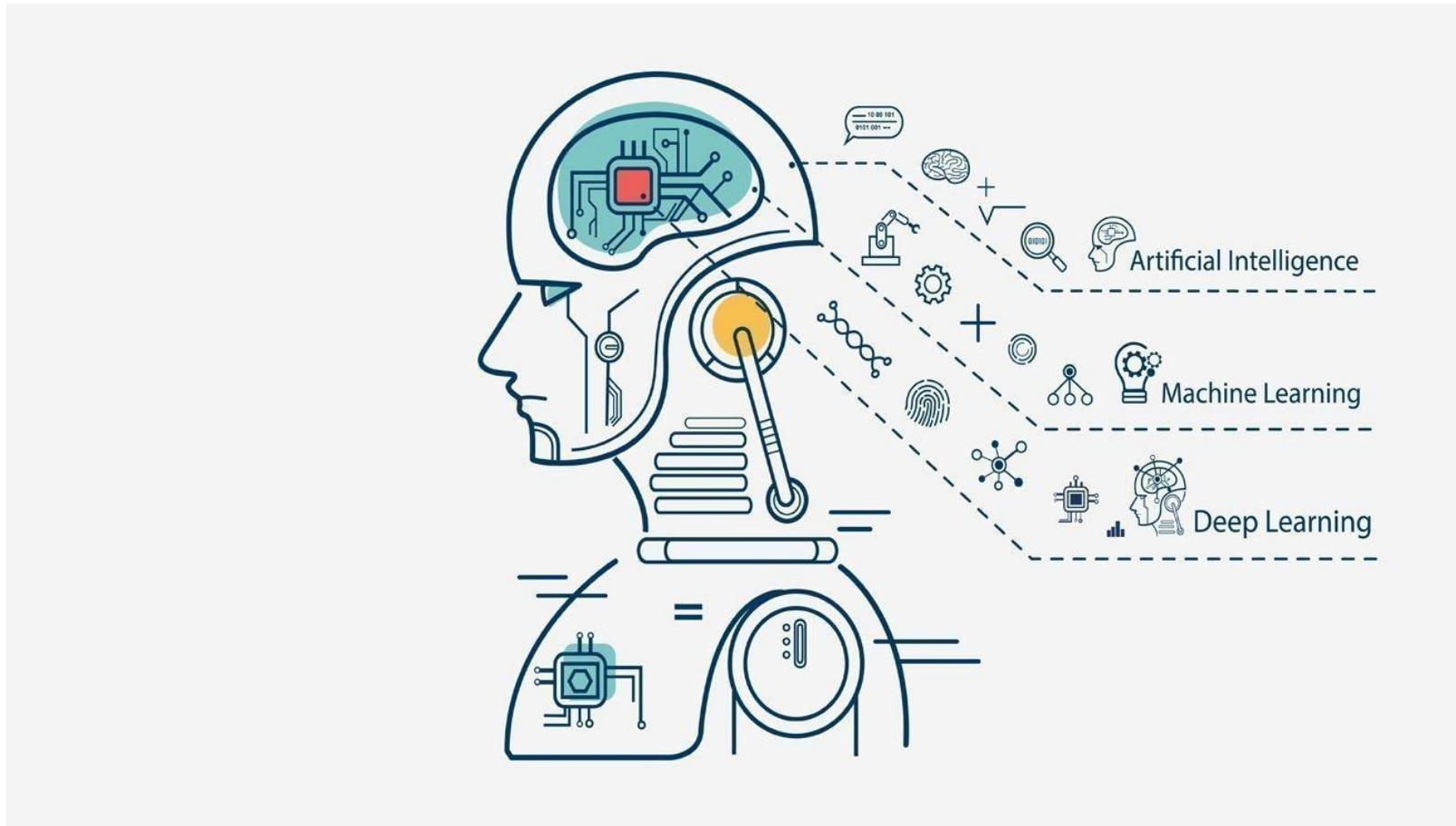
Conocer el contexto del área de aprendizaje máquina y su aplicación en diferentes áreas de estudio.



# Contenido

- Introducción
- Extracción del conocimiento
- Metodología en la minería de datos
- Enfoques de aprendizaje máquina
- Software para Aprendizaje Máquina

# ¿Una máquina puede aprender?



# Definiciones de inteligencia artificial

Sistemas que piensan como humanos	Sistemas que piensan racionalmente
<p>&lt;&lt;El nuevo y excitante esfuerzo de hacer que las computadoras piense... máquinas con mentes, en el más sentido literal&gt;&gt; (Haugeland, 1985)</p> <p>&lt;&lt;[La automatización de] actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje...&gt;&gt; (Bellman, 1978)</p>	<p>&lt;&lt;El estudio de las facultades mentales mediante el uso de modelos computacionales&gt;&gt; (Charniak y McDermott, 1985)</p> <p>&lt;&lt;El estudio de los cálculos que hacen posible percibir, razonar y actuar&gt;&gt; (Winston, 1992)</p>
Sistemas que actúan como humanos	Sistemas que actúan racionalmente
<p>&lt;&lt;El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia&gt;&gt; (Kurzweil, 1990)</p> <p>&lt;&lt;El estudio de cómo lograr que las computadoras realicen tareas que, por el momento, los humanos hacen mejor&gt;&gt; (Rich y Knight, 1991)</p>	<p>&lt;&lt;La Inteligencia computacional es el estudio del diseño de agentes inteligentes&gt;&gt; (Poole et al. 1998)</p> <p>&lt;&lt;IA... está relacionada con conductas inteligentes en artefactos&gt;&gt; (Nilsson, 1998)</p>

# Características de los sistemas de IA



Tratan con problemas que SON DIFICILES para los seres humanos: jugar ajedrez, resolución de teoremas matemáticos, etc.

Tratan con problemas que SON FACILES para los humanos: hablar, reconocer, entender un discurso, etc.

Generalmente, los problemas “difíciles” para los seres humanos, son más fáciles de resolver por los sistemas de IA

# Algunas aplicaciones de la IA

## Medicina

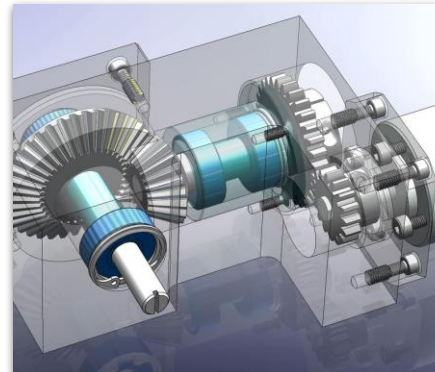


Sistemas expertos de apoyo al diagnóstico médico

Algoritmos Evolutivos aplicados a búsqueda de patrones

Investigación de fármacos

## Ingeniería



Simulación aeroespacial

Robótica y control: autómatas para la optimización de la producción

Redes Neuronales para el control de calidad

## Economía



Predicción de oferta y demanda

Optimización y predicción de mercados

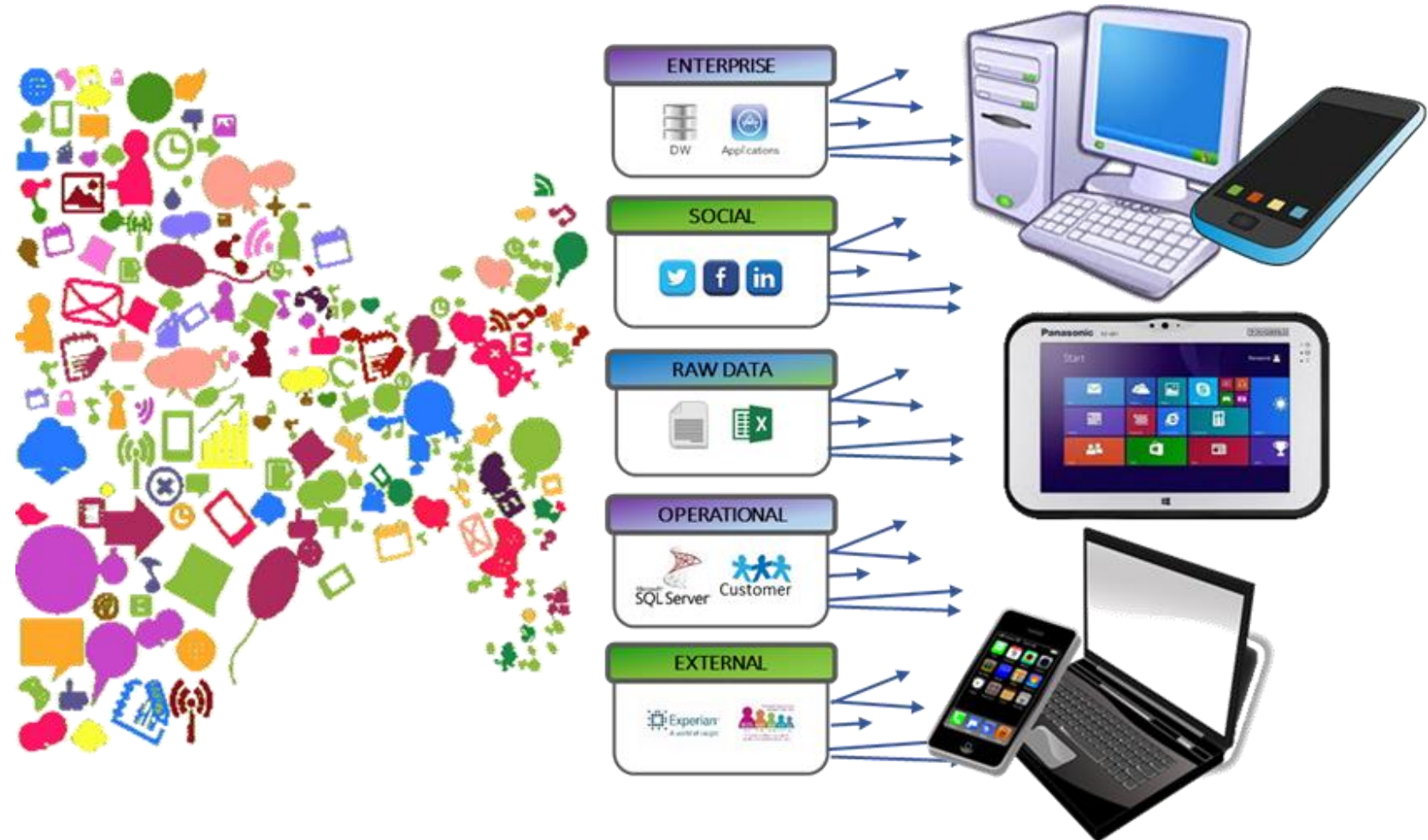
Predicción de riesgos

Descubrimiento de conocimiento en consumo y gastos



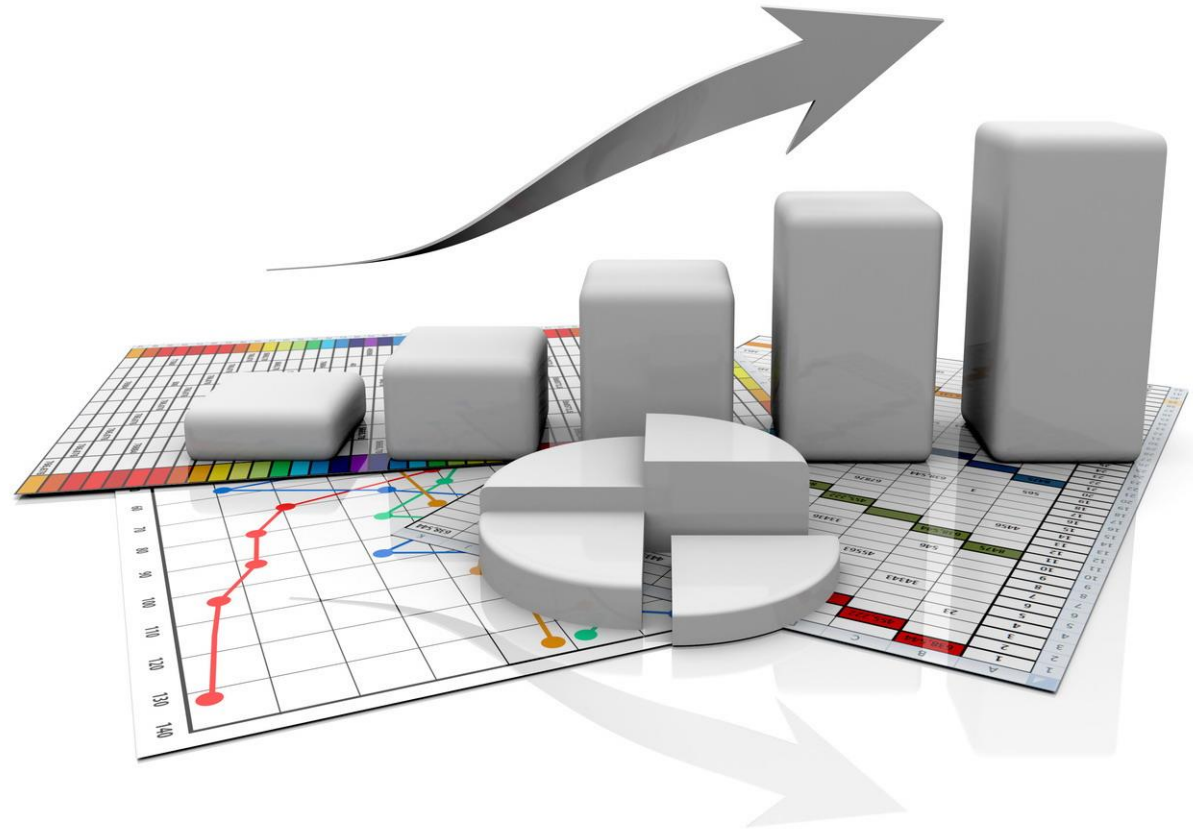
# La explosión de datos

Los sistemas informáticos modernos están acumulando datos a una velocidad casi inimaginable y de una variedad muy amplia de fuentes





dichos datos  
contienen  
información oculta  
que puede ser  
crítica para el  
crecimiento o el  
declive de una  
empresa



## conocimiento que podría:

- conducir a importantes descubrimientos en la ciencia,
- permitir predecir con precisión el clima y los desastres naturales,
- permitir identificar las causas y posibles curas para enfermedades letales,
- significar literalmente la diferencia entre la vida y la muerte.

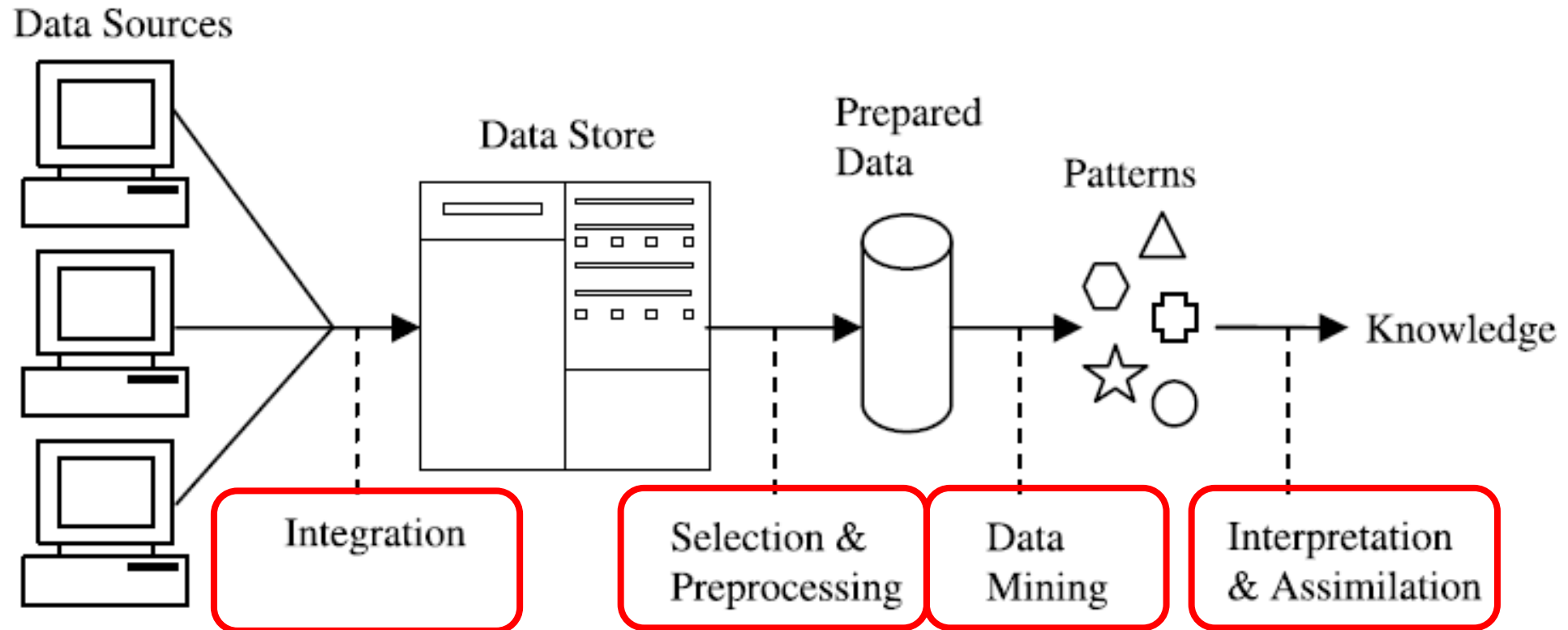


# Análisis de datos

- Es el estudio exhaustivo de un conjunto de información cuyo objetivo es obtener conclusiones que permitan a una empresa o entidad tomar una decisión



# Descubrimiento del Conocimiento en bases de datos



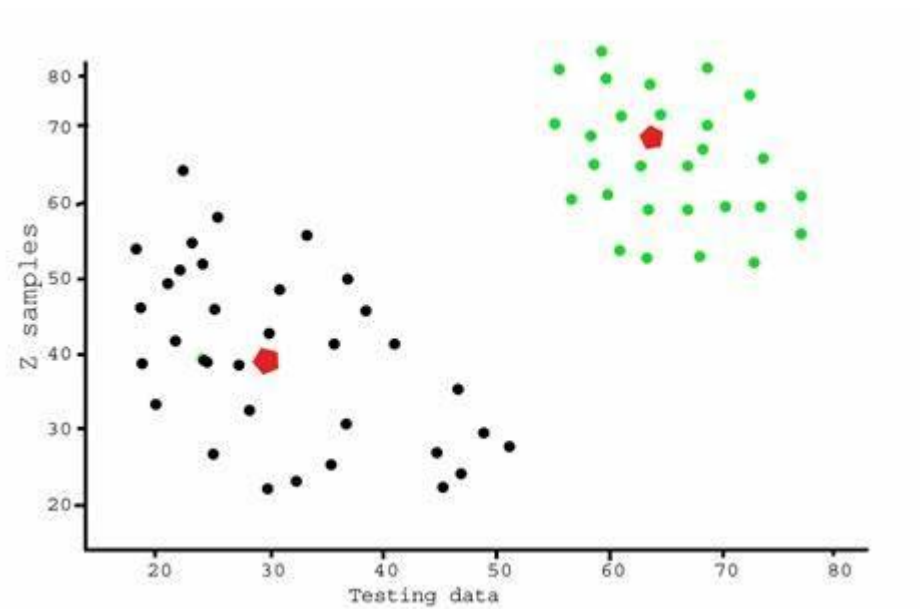
# Minería de datos

- **Data mining** se trata de resolver problemas mediante el análisis de datos ya presentes en las bases de datos



# Minería de datos

- Se define como el **proceso de descubrir patrones en los datos**.
  - El proceso puede ser automático
  - (más usualmente) semiautomático.





# Machine Learning

- ¿Qué es aprender?
- ¿Qué es el aprendizaje automático?





“Aprender” se define como:

- es la **acción de adquirir y retener conocimiento, habilidad o información sobre alguna materia**
- significa adquirir conocimiento a través del intelecto o de la experiencia.
  - El aprendizaje a través del intelecto es dominado por el racionalismo.
  - El conocimiento a través de la experiencia o empirismo, requiere de práctica y mientras más se practica, más se aprende.

# ¿Qué es el aprendizaje máquina?

El aprendizaje automático (ML) es un subcampo de la informática que se ocupa de construir algoritmos que, para ser útiles, se basan en una colección de muestras de algún fenómeno.

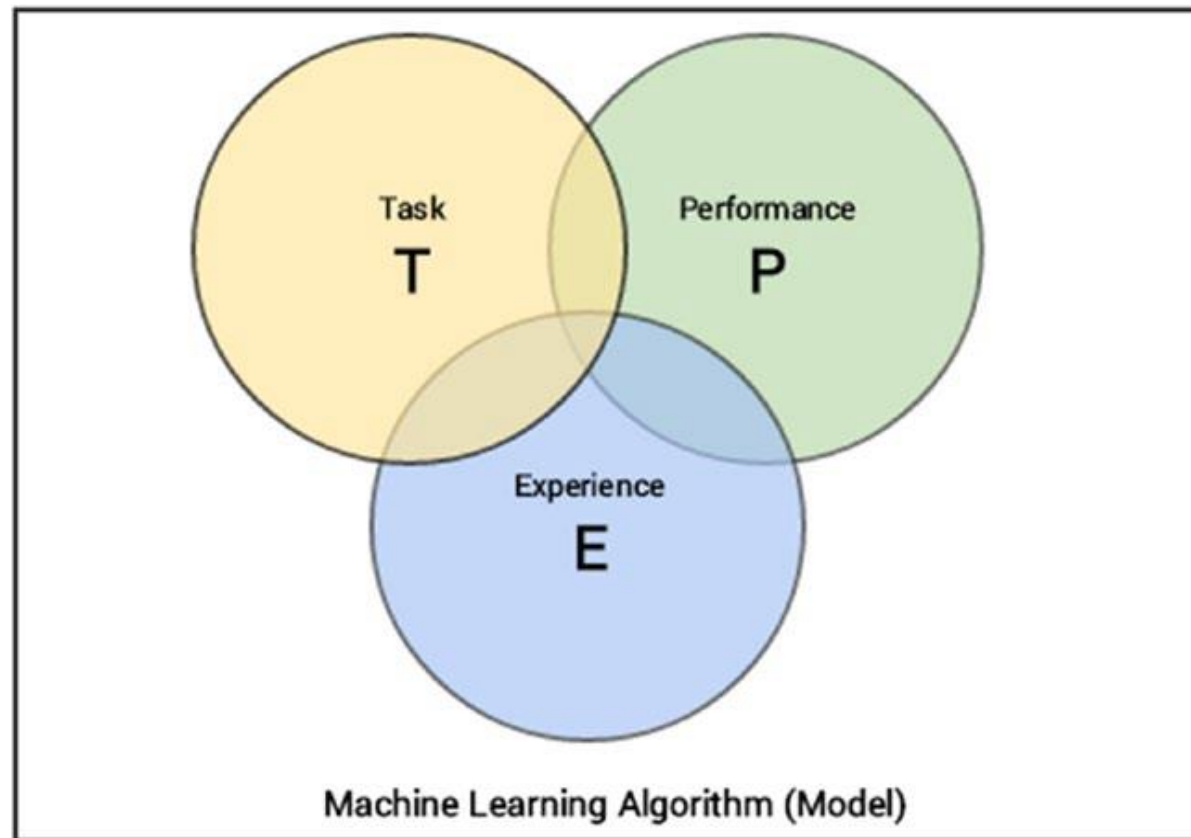
Estas muestras pueden provenir de la naturaleza, ser hechos a mano por humanos o generados por otro algoritmo.



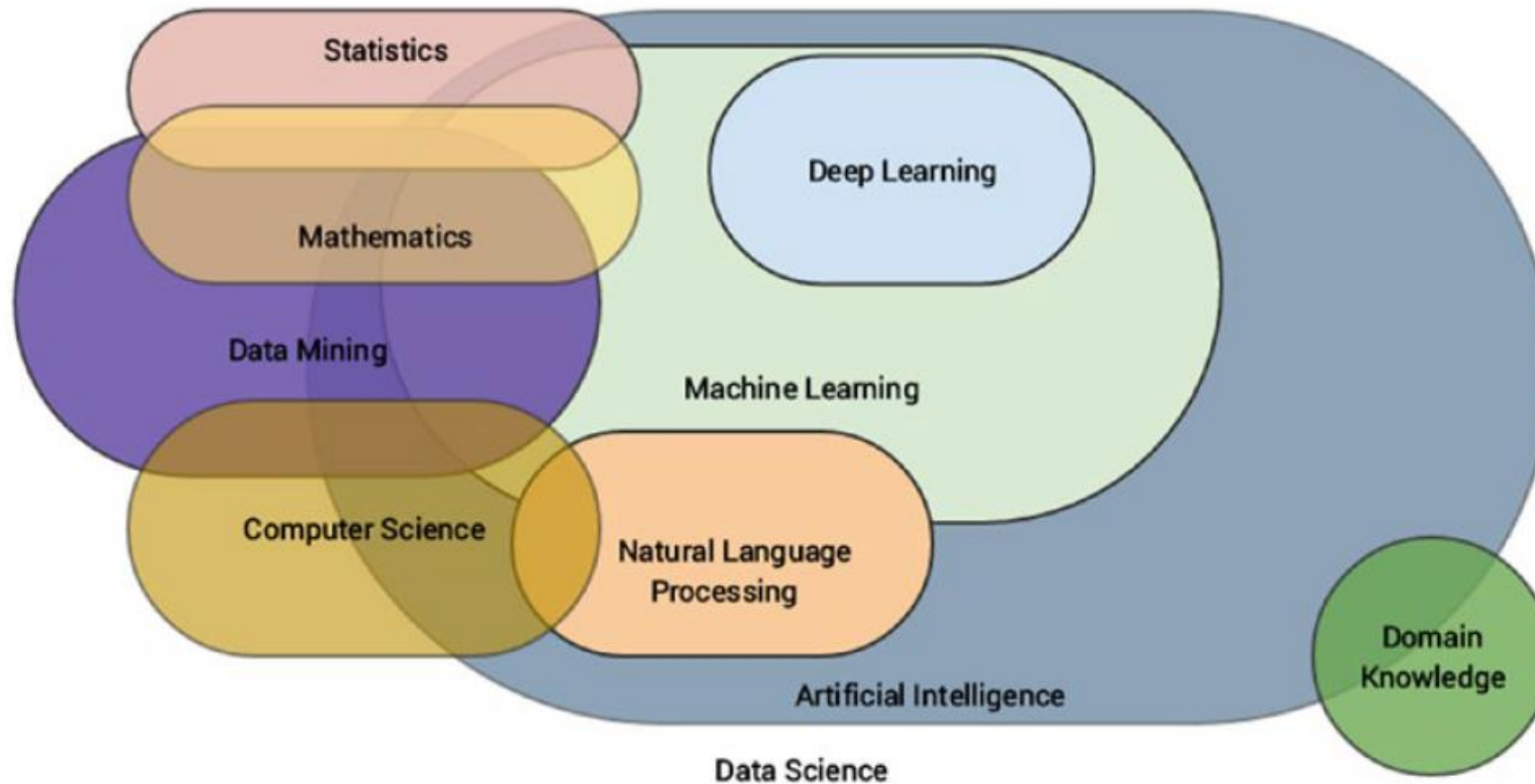
## ¿Por qué hacer que las máquinas aprendan?

- Falta de suficiente experiencia humana en un dominio (por ejemplo, simular navegaciones en territorios desconocidos o incluso planetas espaciales)
- Los escenarios y el comportamiento pueden seguir cambiando con el tiempo (por ejemplo, disponibilidad de infraestructura en una organización, conectividad de red, etc.)
- Los seres humanos tienen suficiente experiencia en el campo, pero es extremadamente difícil explicar formalmente o traducir esta experiencia en tareas computacionales (por ejemplo, reconocimiento de voz, traducción, reconocimiento de escenas, tareas cognitivas, etc.)
- Abordar problemas específicos del dominio a escala con grandes volúmenes de datos con demasiadas condiciones y restricciones complejas.

# Componentes de un algoritmo de aprendizaje



# Aprendizaje automático: un verdadero campo multidisciplinario



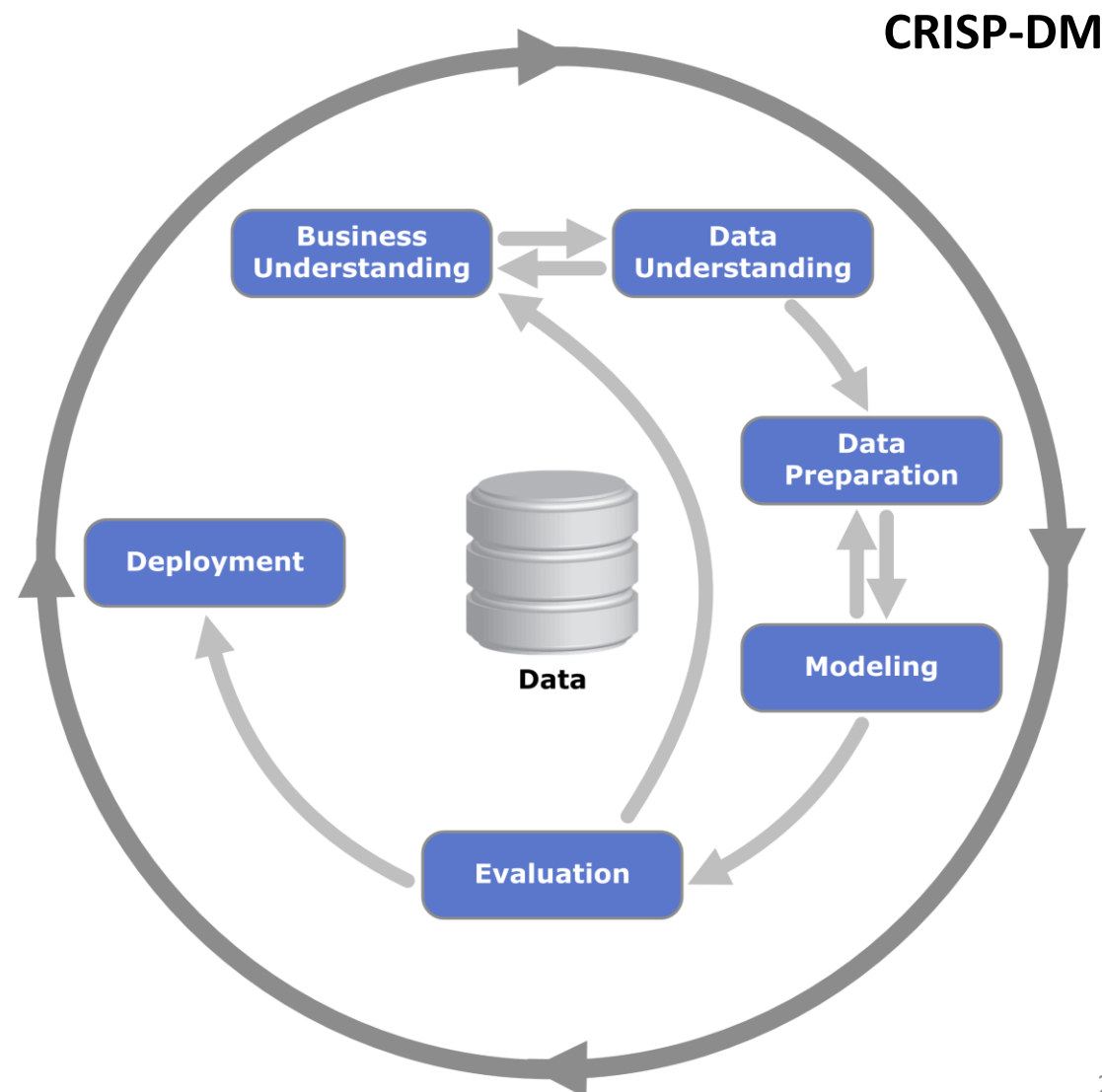
# Aplicaciones

- análisis de compuestos orgánicos
- abstracción automática
- detección de fraude con tarjeta de crédito
- predicción de carga eléctrica
- previsión financiera
- diagnostico medico
- predicción de la participación del público televisivo
- diseño de producto
- valoración inmobiliaria
- marketing dirigido
- resumen del texto
- optimización de la central térmica
- análisis de peligros tóxicos
- predicción del clima

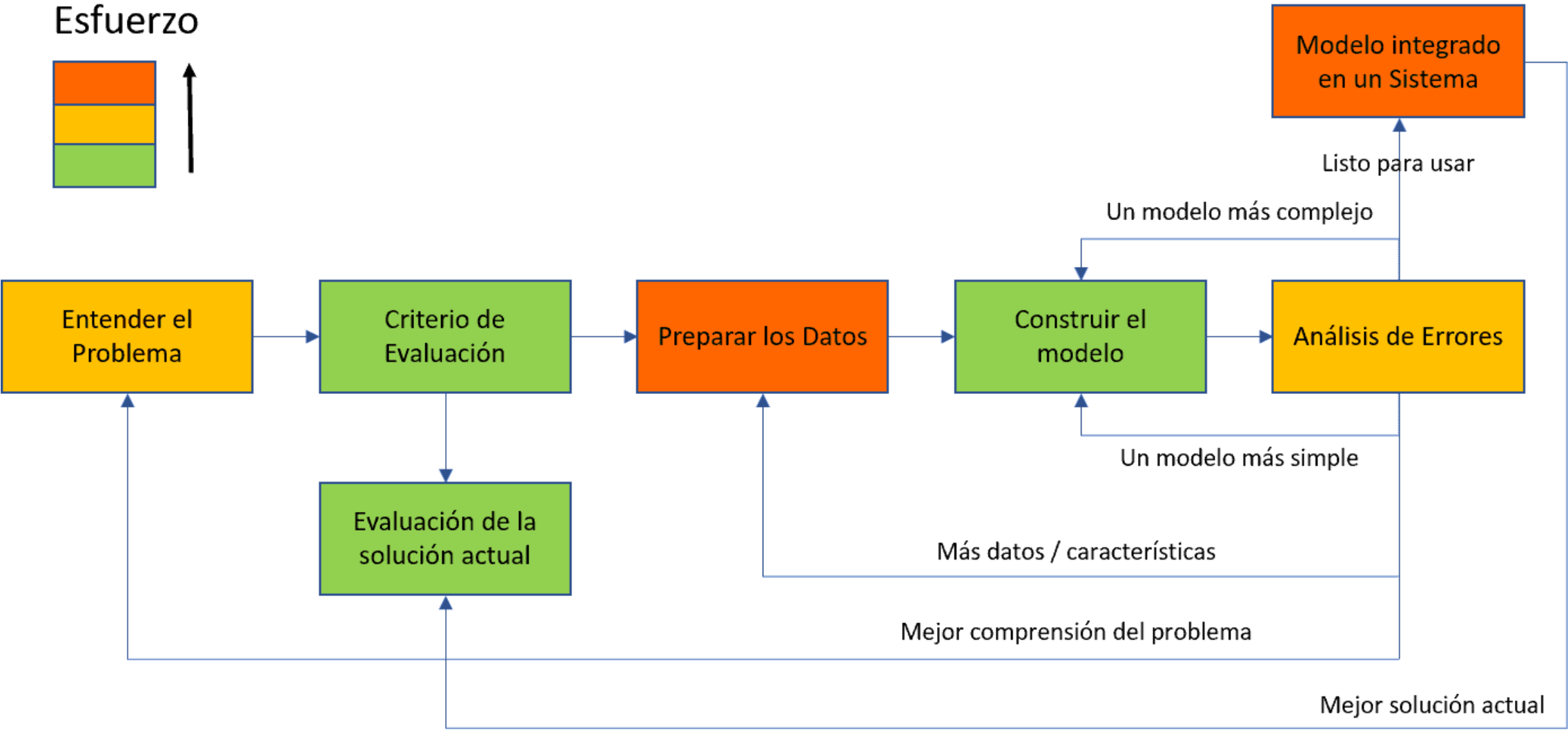


# El proceso de minería de datos

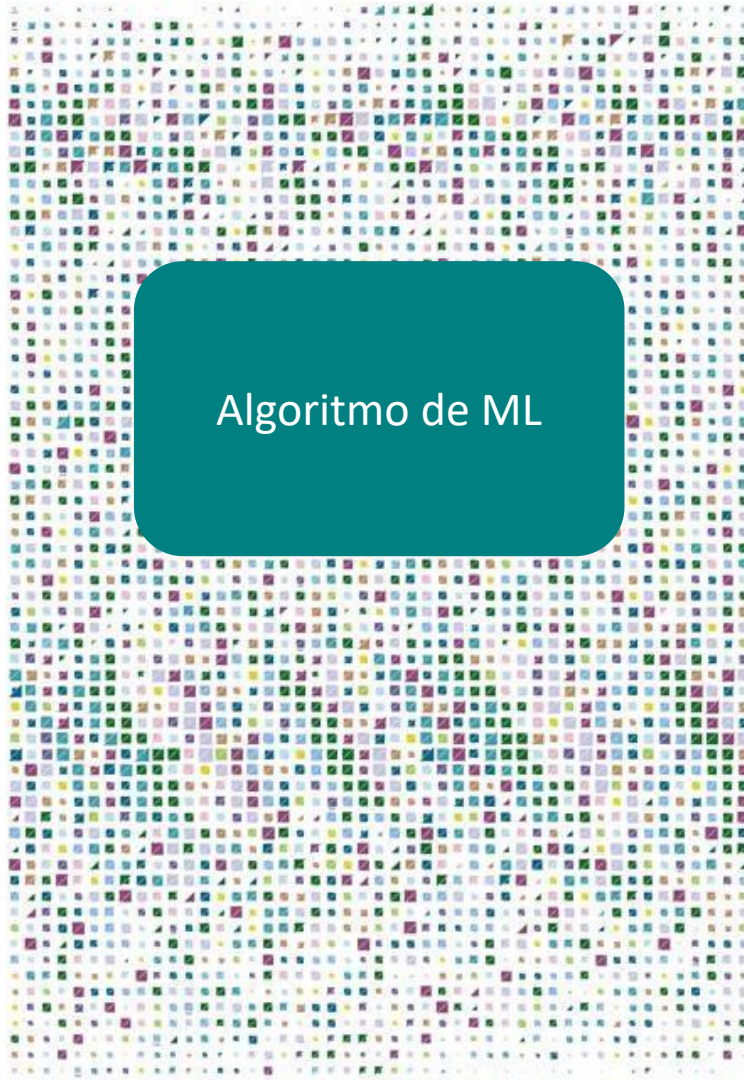
- Cross Industry Standard Process for Data Mining)
  - Es un modelo estándar abierto del proceso que describe los enfoques comunes que utilizan los expertos en minería de datos.
  - Es el modelo analítico más usado



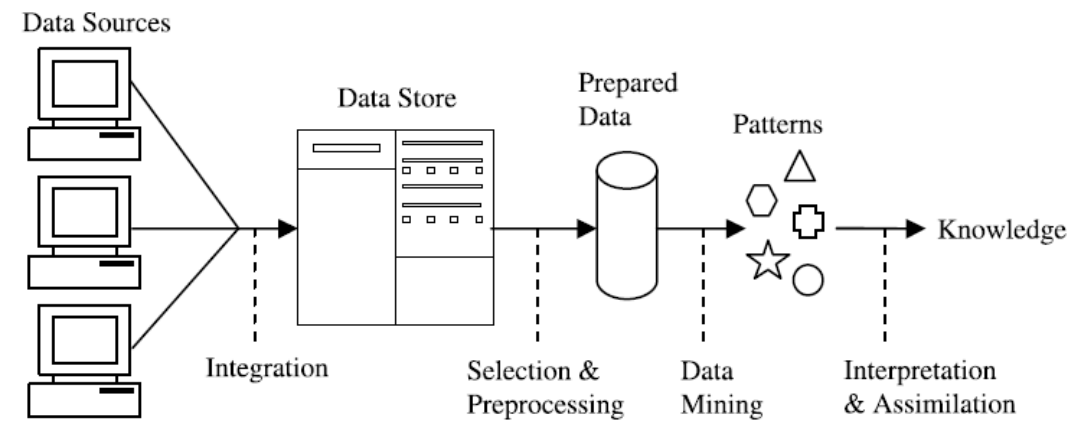




# Patrones en datos



¿Y cómo se expresan los patrones?



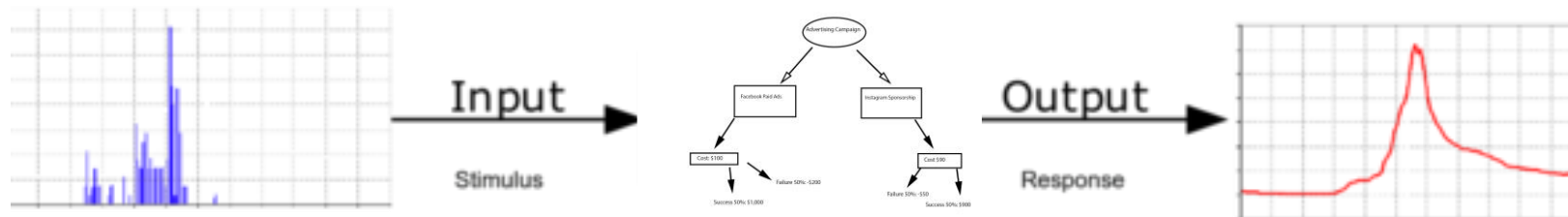
# Patrones de caja negra

1. como una caja negra cuyas entrañas son efectivamente incomprensibles



# Patrones estructurales

2. como una caja transparente cuya construcción revela la estructura del patrón.



# Diferencias

---

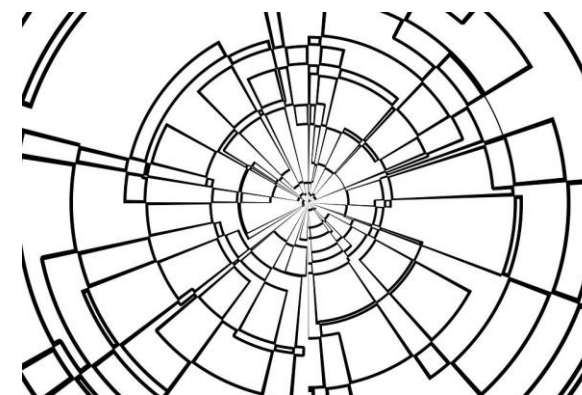
La diferencia entre ambas expresiones es: **si los patrones que se extraen están representados o no** en términos de una estructura que puede ser examinada, razonada y utilizada para informar futuras decisiones.

---

Los llamamos estructurales capturan la estructura de decisión de manera explícita. En otras palabras, ayudan a explicar algo sobre los datos.

# Describiendo patrones estructurales

- ¿Qué se entiende por patrones estructurales?
- ¿Cómo los describes?
- ¿qué forma toma la entrada?



Responderemos estas preguntas a modo de ilustración en lugar de intentar definiciones formales

# Lentes de contacto

	Edad	Prescripción de anteojos	Astigmatismo	Tasa de producción de lágrimas	Lentes recomendadas
1	joven	miope	no	reducida	no
2	joven	miope	no	normal	suave
3	joven	miope	si	reducida	no
4	joven	miope	si	normal	duro
5	joven	hipermetrope	no	reducida	no
6	joven	hipermetrope	no	normal	suave
7	joven	hipermetrope	si	reducida	no
8	joven	hipermetrope	si	normal	duro
9	pre-presbicia	miope	no	reducida	no
10	pre-presbicia	miope	no	normal	suave
11	pre-presbicia	miope	si	reducida	no
12	pre-presbicia	miope	si	normal	duro
13	pre-presbicia	hipermetrope	no	reducida	no
14	pre-presbicia	hipermetrope	no	normal	suave
15	pre-presbicia	hipermetrope	si	reducida	no
16	pre-presbicia	hipermetrope	si	normal	no
17	presbicia	miope	no	reducida	no
18	presbicia	miope	no	normal	no
19	presbicia	miope	si	reducida	no
20	presbicia	miope	si	normal	duro
21	presbicia	hipermetrope	no	reducida	no
22	presbicia	hipermetrope	no	normal	suave
23	presbicia	hipermetrope	si	reducida	no
24	presbicia	hipermetrope	si	normal	no

- 1. edad del paciente: (1) joven, (2) pre-presbicia, (3) presbicia
- 2. prescripción de gafas: (1) miope, (2) hipermetrope
- 3. astigmático: (1) no, (2) sí
- 4. tasa de producción de lágrimas: (1) reducida, (2) normal

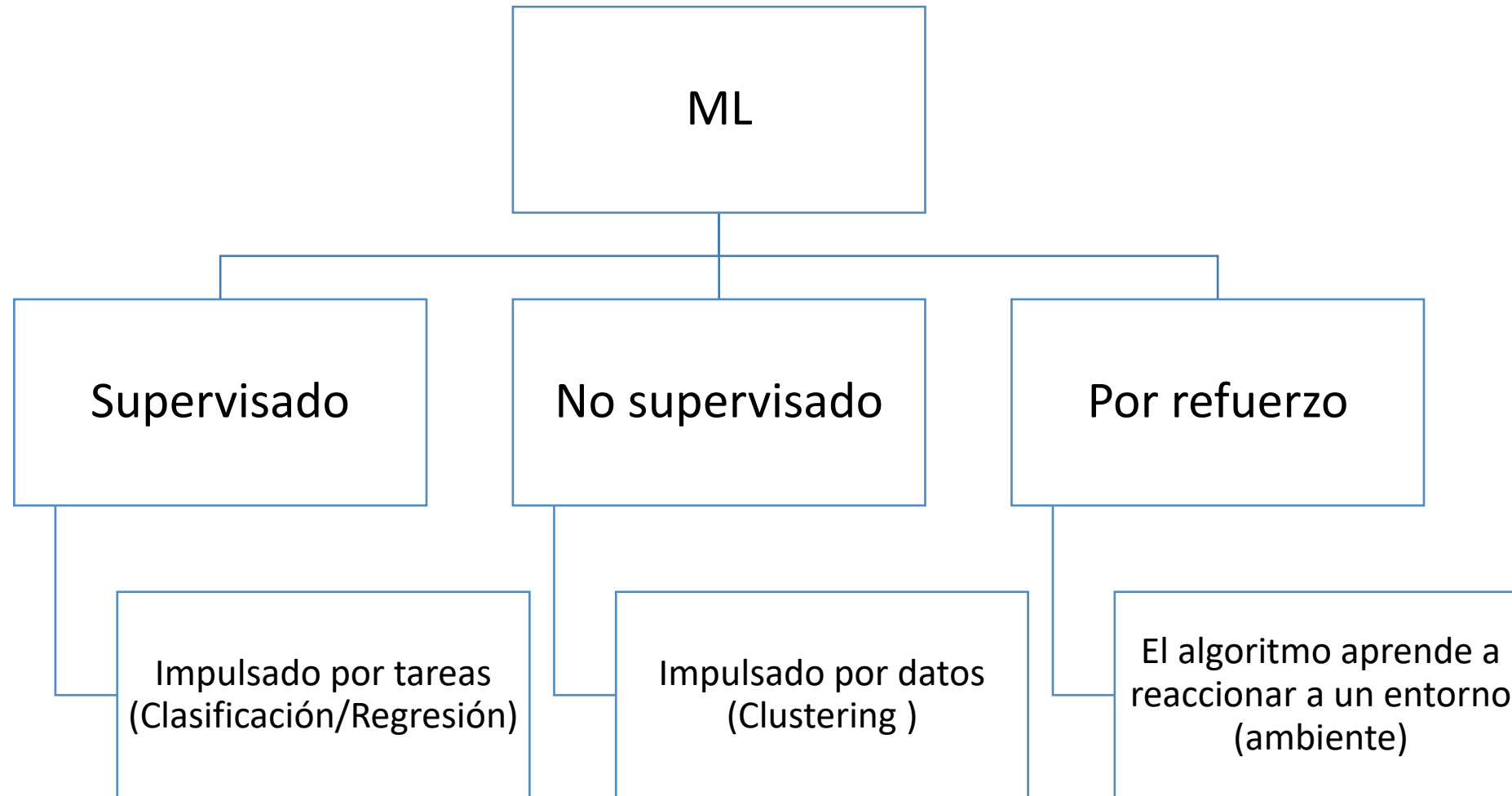


# Estructura

- Regla

SI tasa\_producción\_lágrimas = reducida **ENTONCES** recomendación=no  
EN OTRO CASO, SI edad=joven Y astigmatico=no **ENTONCES** recomendacion=suave

# Técnicas de ML



# Weka

- **Weka** (**W**aikato **E**nvironment for **K**nowledge **A**nalysis)
- Escrito en Java desde 1997
- software libre (licencia GNU-GPL)
- La primera versión en 1993 (TCL/Tk y C)



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*



- En 2005 fue premiado con el ACM SIGKDD Service Award.
- reconoce a un individuo o un grupo por sus destacadas contribuciones a los servicios profesionales en el campo del descubrimiento de conocimiento y la minería de datos



<https://www.kdd.org/awards/view/2005-sigkdd-service-award-the-weka-team>

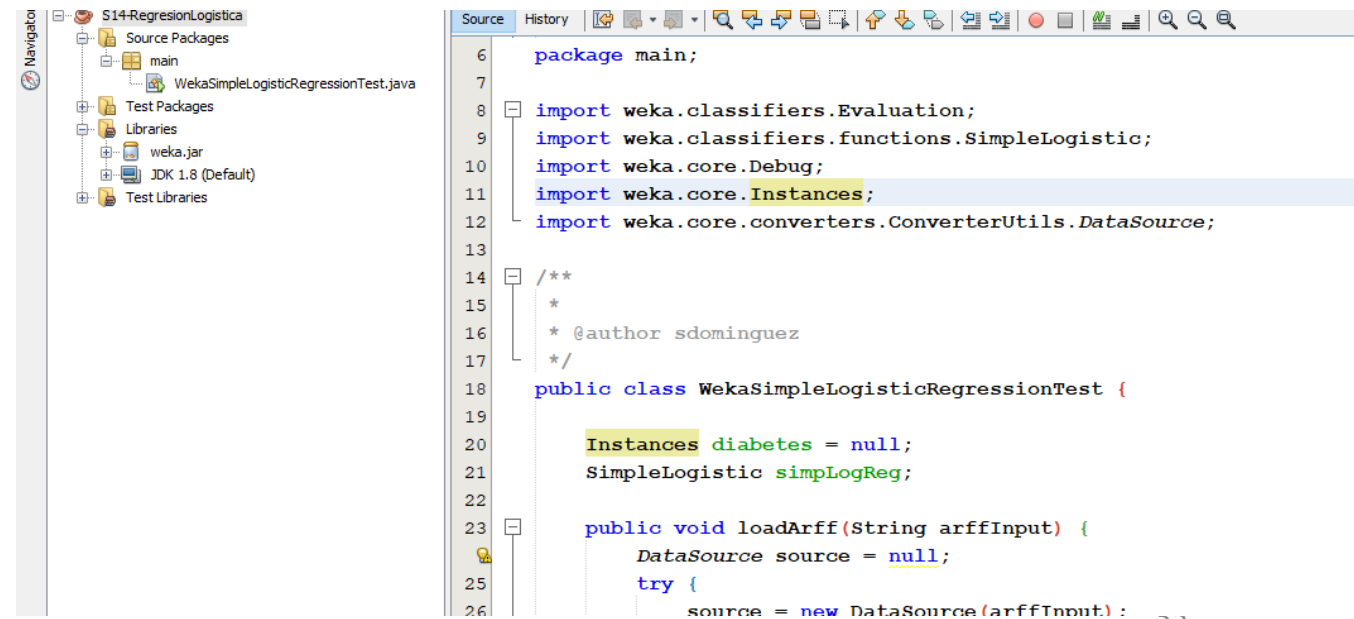
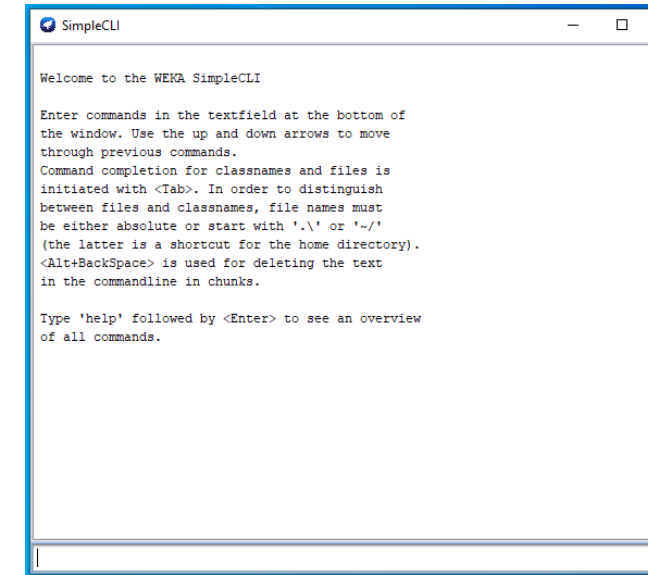
- En 2006 Pentaho Corporation adquirió una licencia para usar Weka para Inteligencia de negocio (Business Intelligence),
  - originando el componente de minería de datos y análisis predictivo del paquete de software Pentaho Business Intelligence.



# WEKA

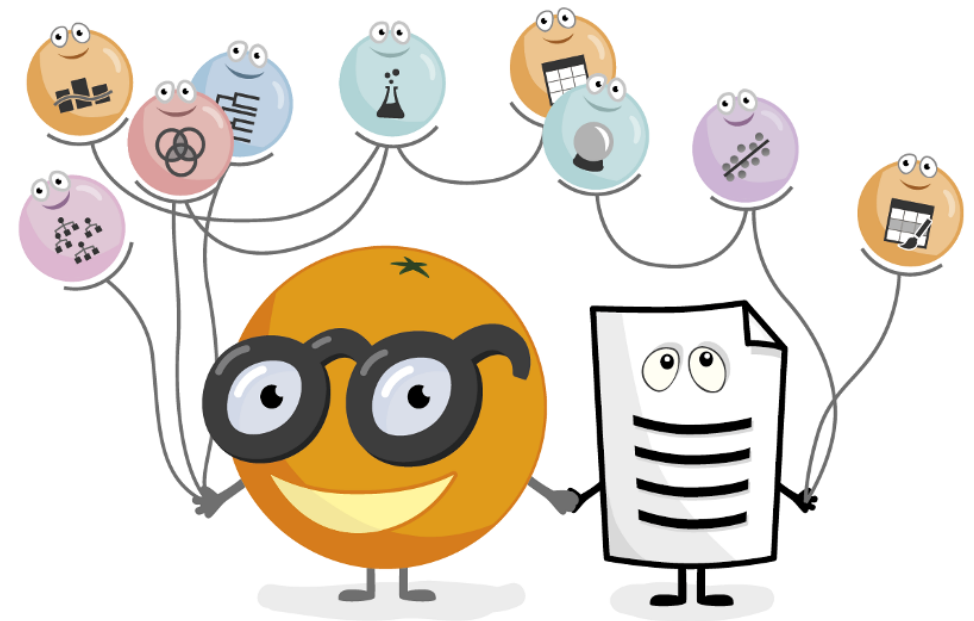
Tiene diferentes opciones de uso

<https://www.cs.waikato.ac.nz/ml/weka/>



# Orange Data Mining

- Herramienta para realizar minería de datos y análisis predictivo.
- Consta de una serie de componentes desarrollados en C++ que implementan algoritmos de minería de datos, así como operaciones de preprocesamiento y representación gráfica de datos.



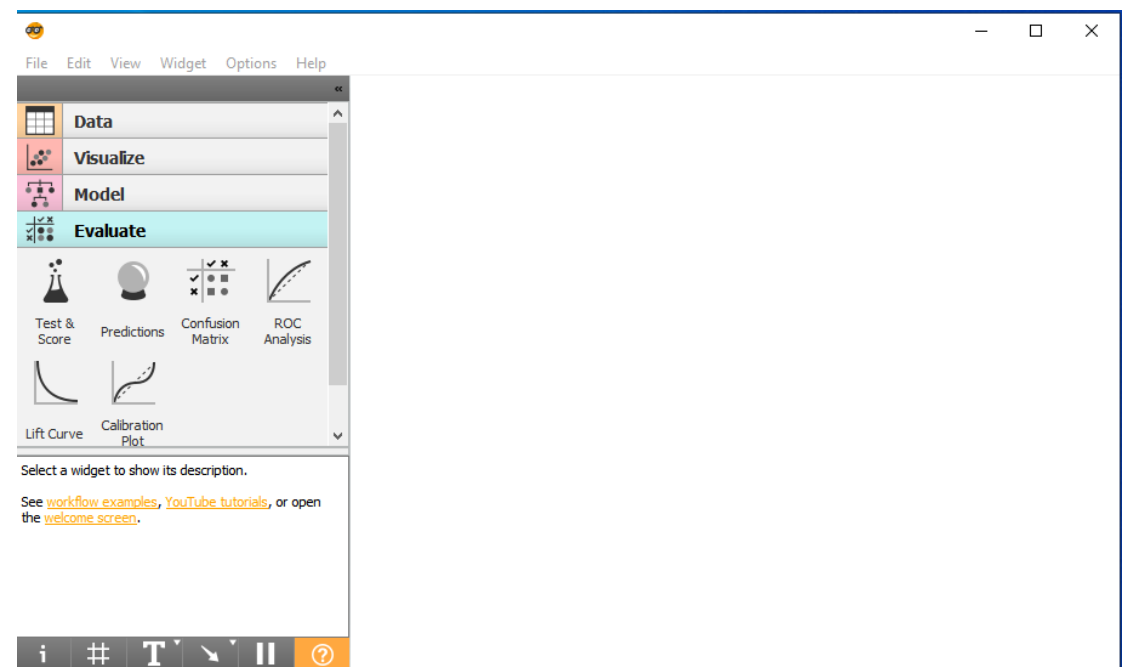


- desarrollado por el laboratorio de Bioinformática de la Universidad de Ljubljana
- Se distribuye bajo licencia GPL.



- Los componentes de Orange pueden ser manipulados desde programas desarrollados en Python o a través de un entorno gráfico (Canvas).

```
% python  
>>> import Orange  
>>> Orange.version.version  
'3.25.0.dev0+3bdef92'  
>>>
```



<https://orange.biolab.si/>

