

Taller: Introducción al Procesamiento de Lenguaje Natural

Jornada Académica NovaUniversitas.

“Innovar para transformar: conexiones que impulsan el desarrollo”

Taller: Introducción al PLN

Objetivo

Este taller está diseñado para estudiantes de primer semestre de carreras en computación o informática interesados en iniciarse en el campo del Procesamiento de Lenguaje Natural (PLN)

Taller: Introducción al PLN

Contenido

1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?
2. Representación de texto
3. Aplicaciones básicas de PLN
4. Reflexiones

Taller: Introducción al Procesamiento de Lenguaje Natural

1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

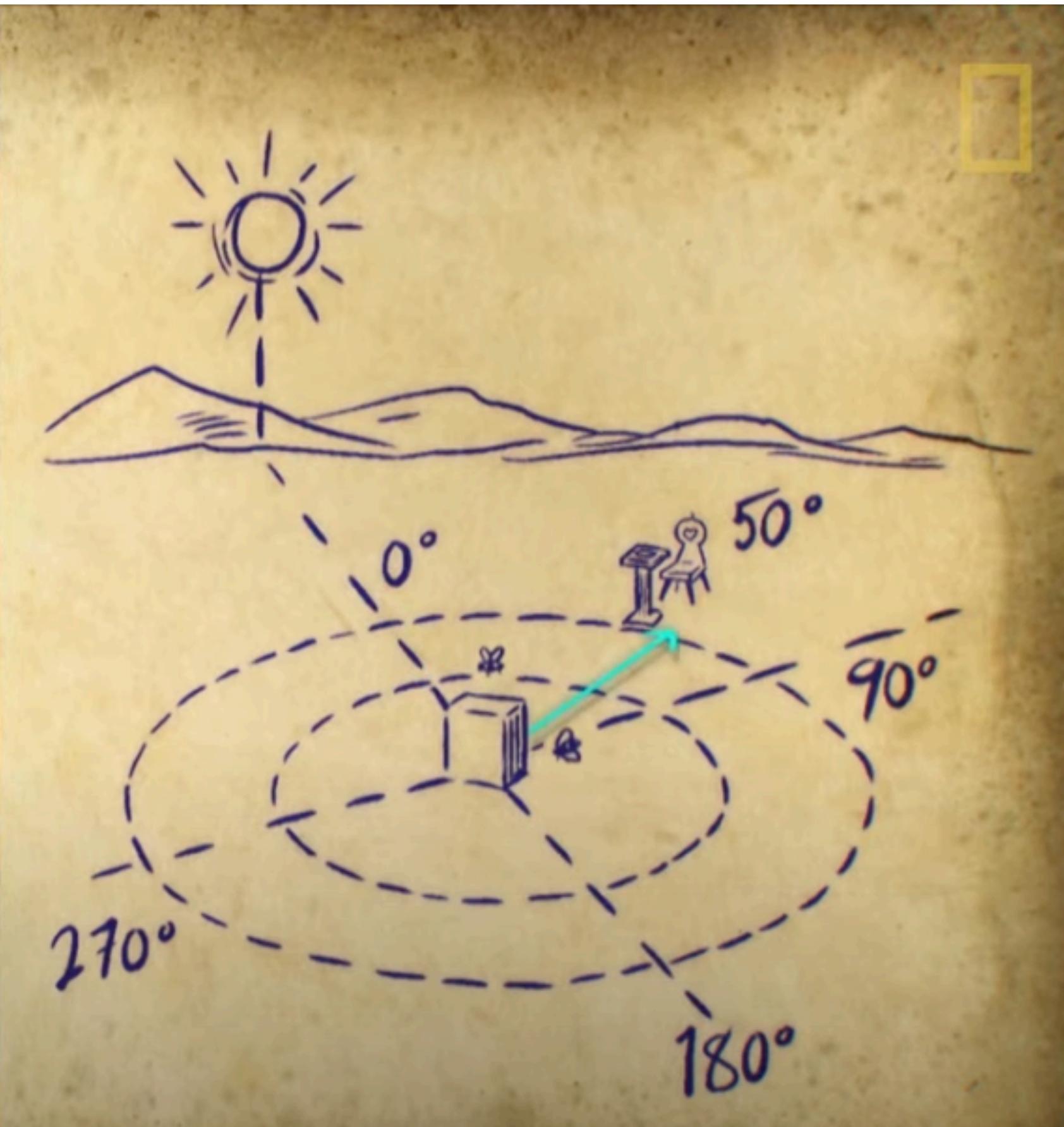
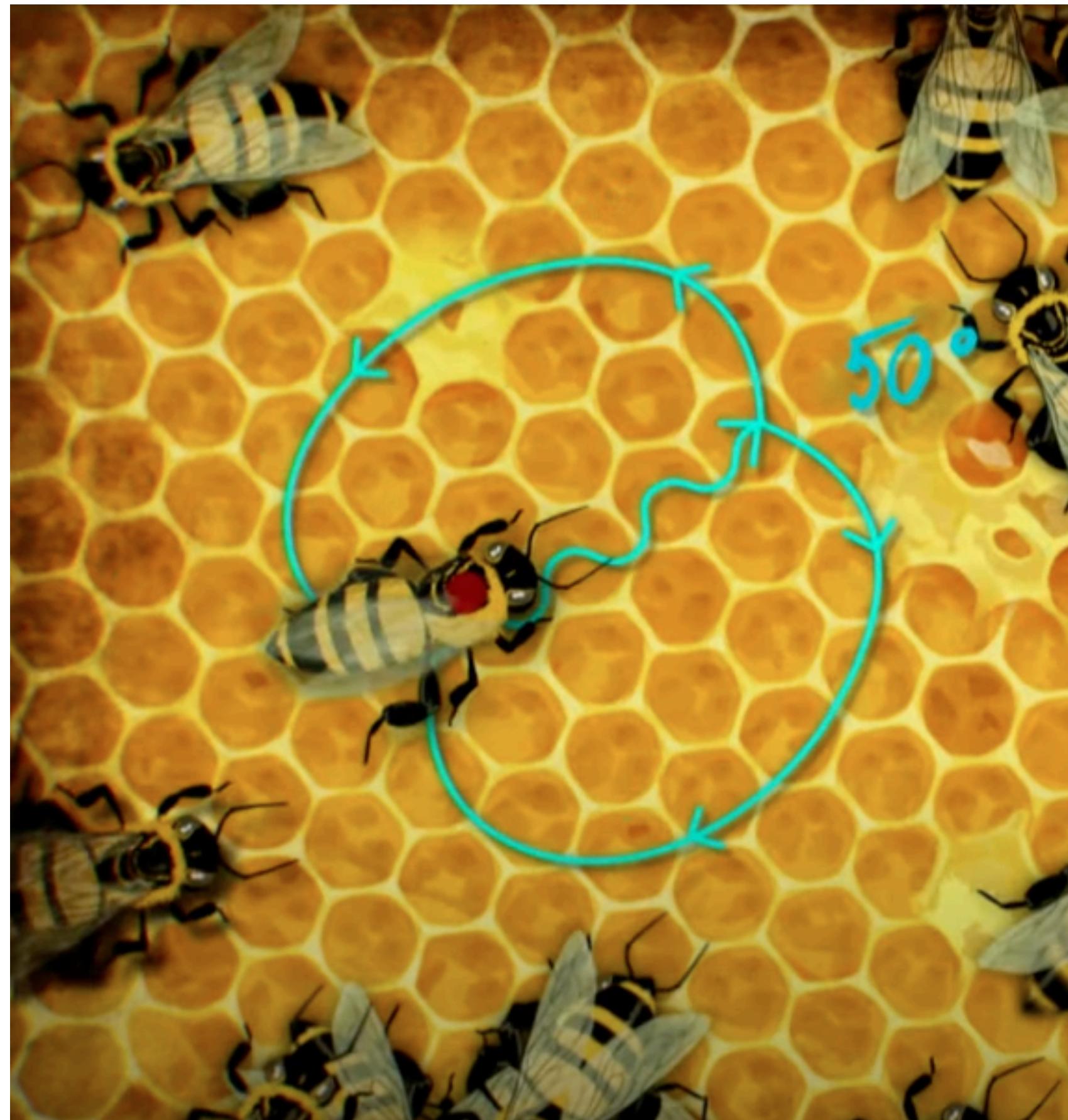
1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

Es un sistema de comunicación



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

Es un sistema de comunicación



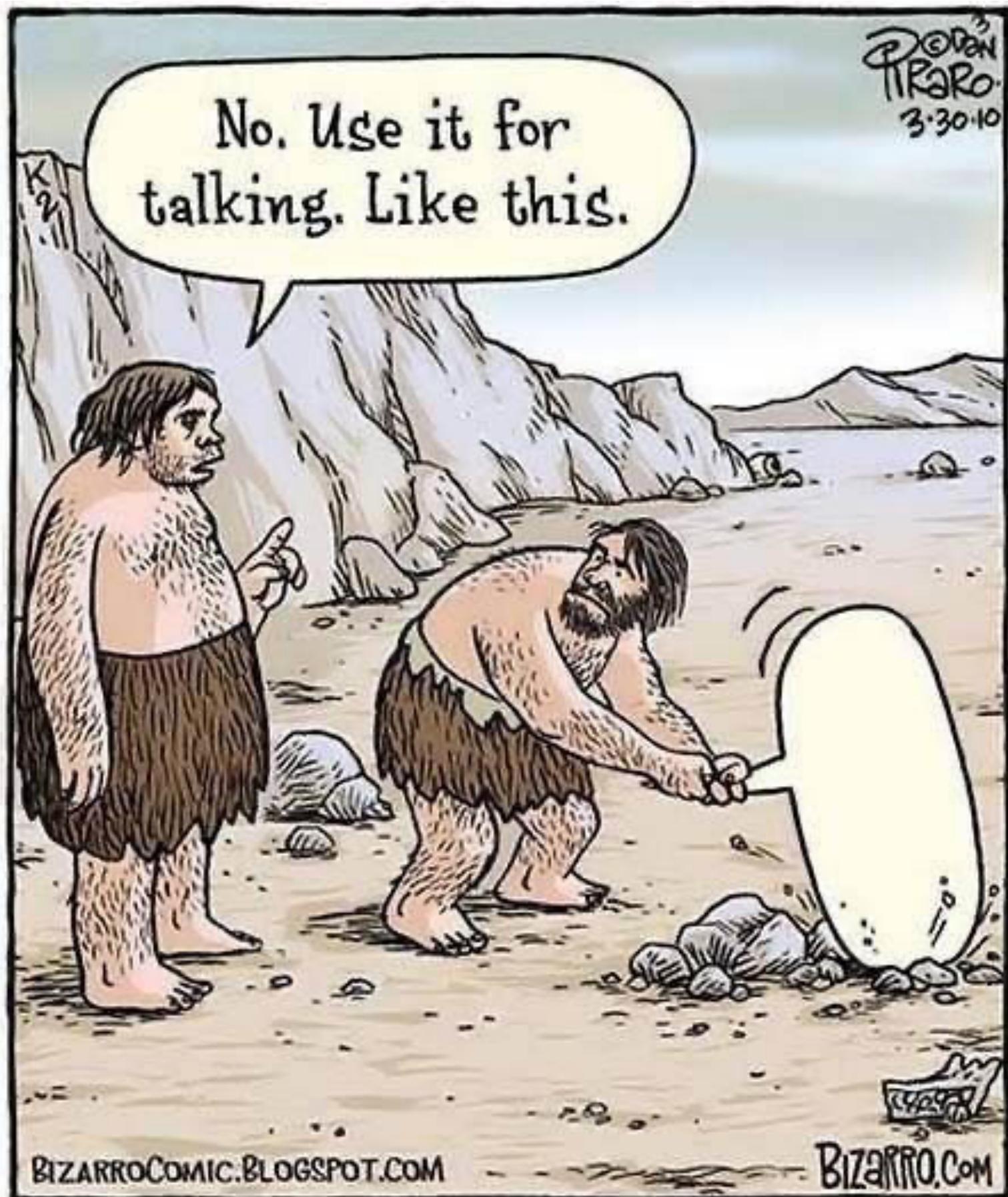
1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

Es un sistema de comunicación



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

Tres dimensiones del lenguaje humano



- Infinidad discreta
- Desplazamiento
- Atención conjunta

1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

El lenguaje humano

El **lenguaje** es, sin duda, una de las **características más distintivas** de los **seres humanos**.

Lo utilizamos para **comunicarnos**, estructurar nuestro **pensamiento** y **expresar nuestra identidad**, mostrando a qué grupo pertenecemos o al que deseamos pertenecer.



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

El lenguaje humano

El aspecto esencial del lenguaje natural es que tiene hablantes nativos



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

El lenguaje humano

Además, permite expresar ideas sobre cualquier tema, sin restricciones.



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

El lenguaje humano

Debe existir una comunidad de hablantes que la reconozca como su lengua nativa y la utilice para la comunicación cotidiana.



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

El lenguaje natural

Un **lenguaje natural** es aquel que ha sido adquirido por una comunidad de personas desde la infancia y ha evolucionado a lo largo de generaciones.



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

Diversidad lingüística

¿Cuántos idiomas existen en el mundo?

La estimación más aceptada indica que hay entre 6000 y 7000 idiomas en el mundo.

How many languages are there in the world?

7,159 languages are in use today. □

That number is constantly in flux, because we're learning more about the world's languages every day. And beyond that, the languages *themselves* are in flux. They're living and dynamic, used by communities whose lives are shaped by our rapidly changing world. This is a fragile time: Roughly 44% of all languages are now endangered, often with fewer than 1,000 users remaining. Meanwhile, the world's 20 largest languages are the native tongue of more than 3.6 billion people total. That's just 0.3% of the world's languages accounting for nearly half of the world's population!

Living Languages, 2022

Map by Ethnologue by SIL International

1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

PLN



“El lenguaje es la joya de la corona de la cognición.”
- Steven Pinker, 1994

1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

PLN

El PLN es una rama de la IA.

1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

PLN

El procesamiento de lenguaje natural (NLP, Natural Language Processing) es una intersección de los campos de la ciencias de la computación, lingüística y inteligencia artificial.

Su objetivo principal es permitir que las máquinas comprendan, interpreten y generen texto de manera similar a como lo hacen los humanos.

1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

PLN

El procesamiento de lenguaje natural (NLP, Natural Language Processing) es una intersección de los campos de la ciencias de la computación, lingüística y inteligencia artificial.

Su objetivo principal es permitir que las máquinas **comprendan** y **generen** texto de manera similar a como lo hacen los humanos.

1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

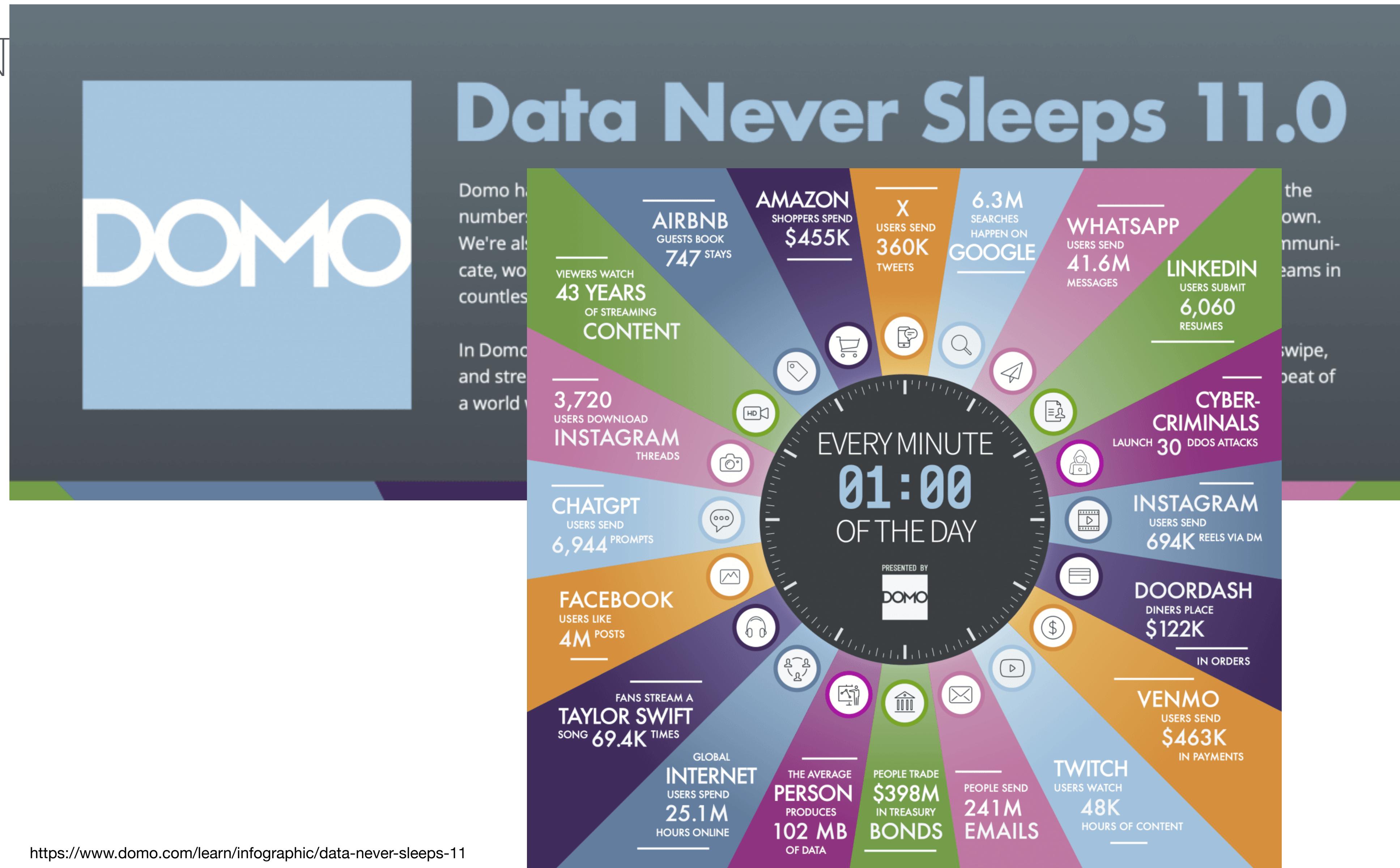
¿Por qué PLN?

El 80% de los datos de las empresas son datos textuales, como registros, mensajes, blogs, artículos, documentos, entre otros.

El PLN ha ganado importancia porque las compañías están invirtiendo en equipos de ciencia de datos para usar los datos con el propósito de extraer los tipos de insights correctos.

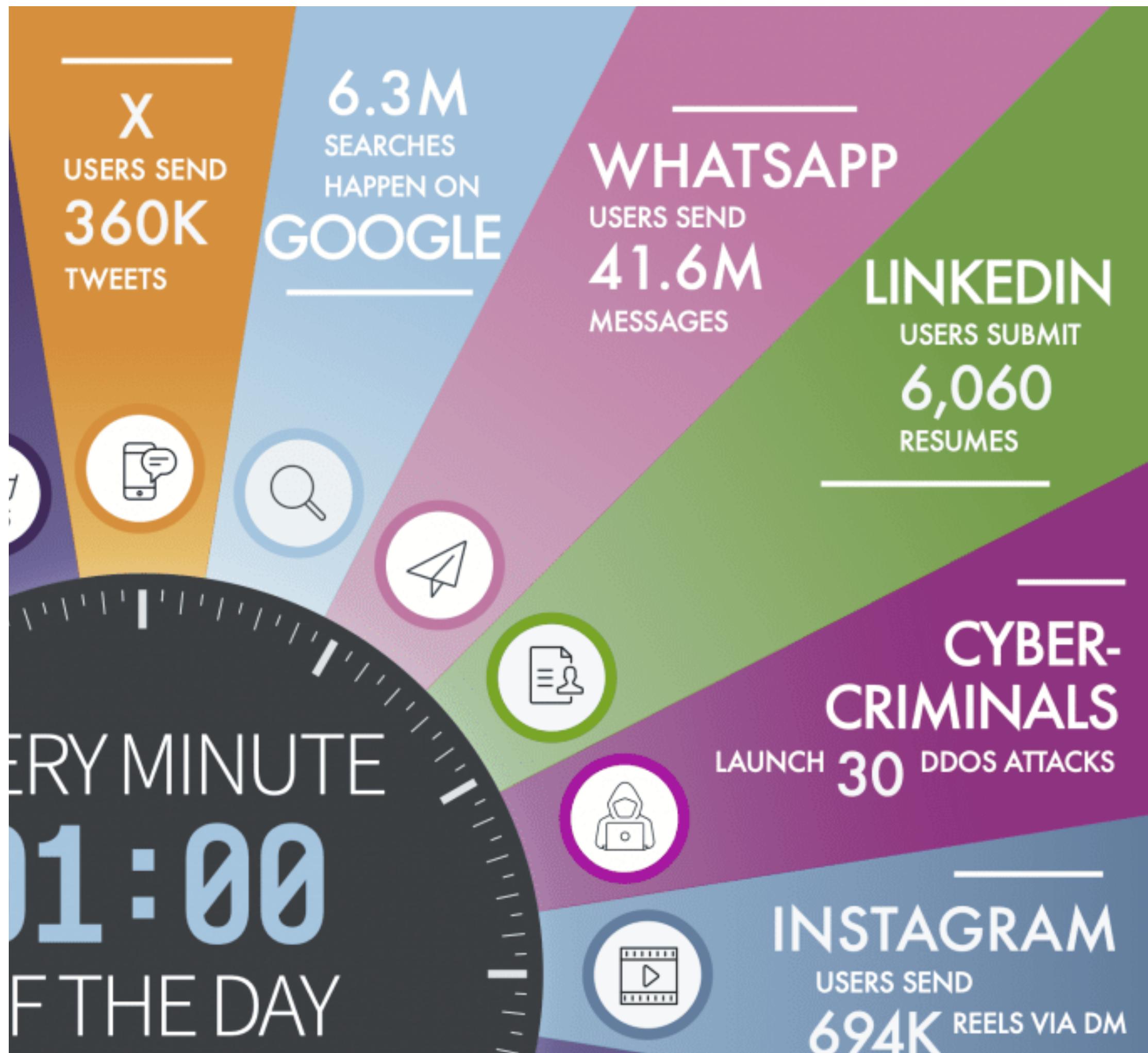
1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

¿Por qué PLN



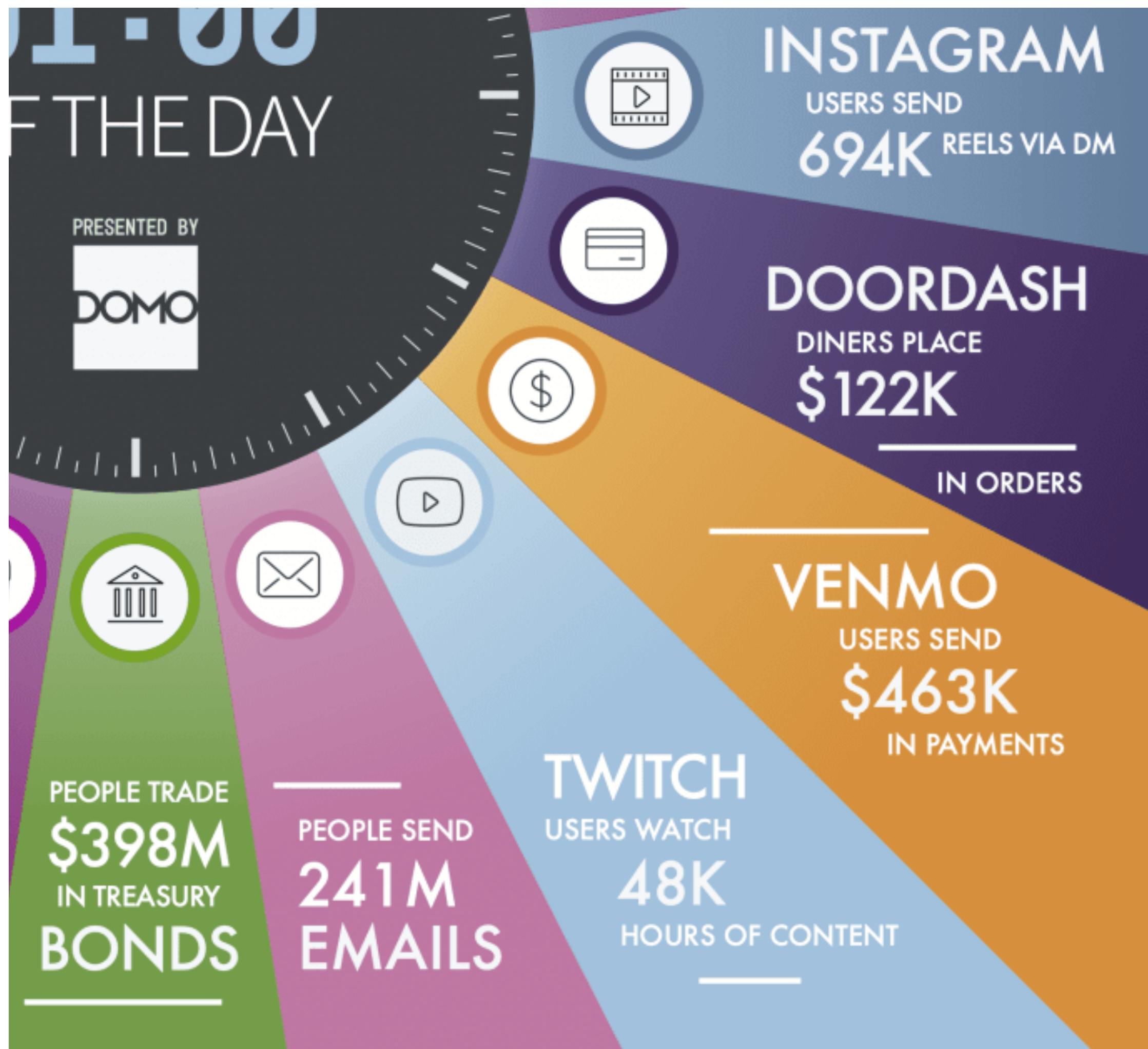
1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

¿Por qué PLN?



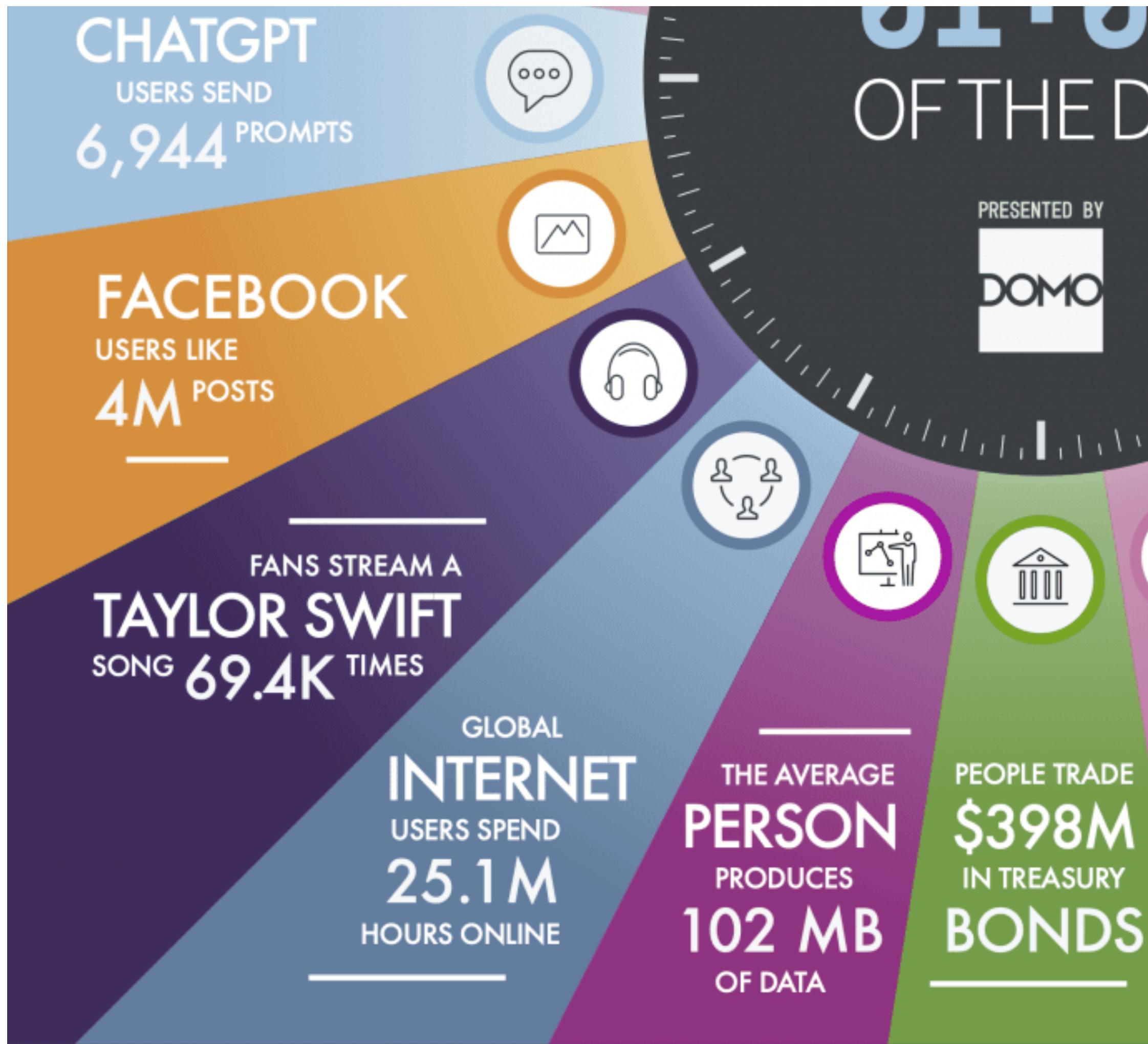
1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

¿Por qué PLN?



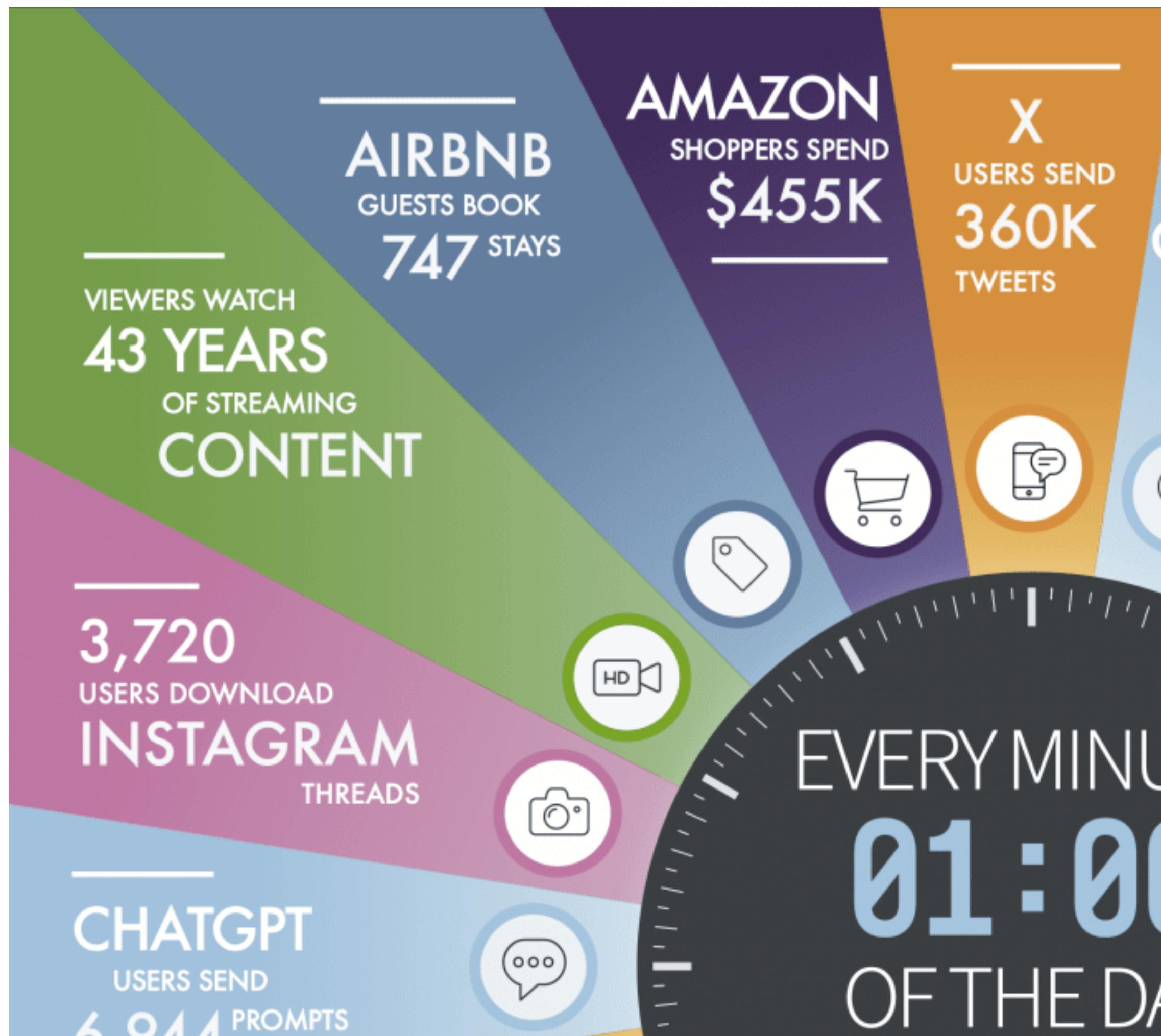
1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

¿Por qué PLN?



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

¿Por qué PLN?



1. El Procesamiento de Lenguaje Natural: ¿Para qué sirve?

Aplicaciones del PLN

Clasificación de texto

Agrupación de texto y similitud

Búsqueda y recuperación de información

Parsing y reconocimiento de entidades nombradas

Resúmenes de textos

Traducción automática

Interfaces conversacionales

- Chatbots

- Asistentes personales

- Q&A systems

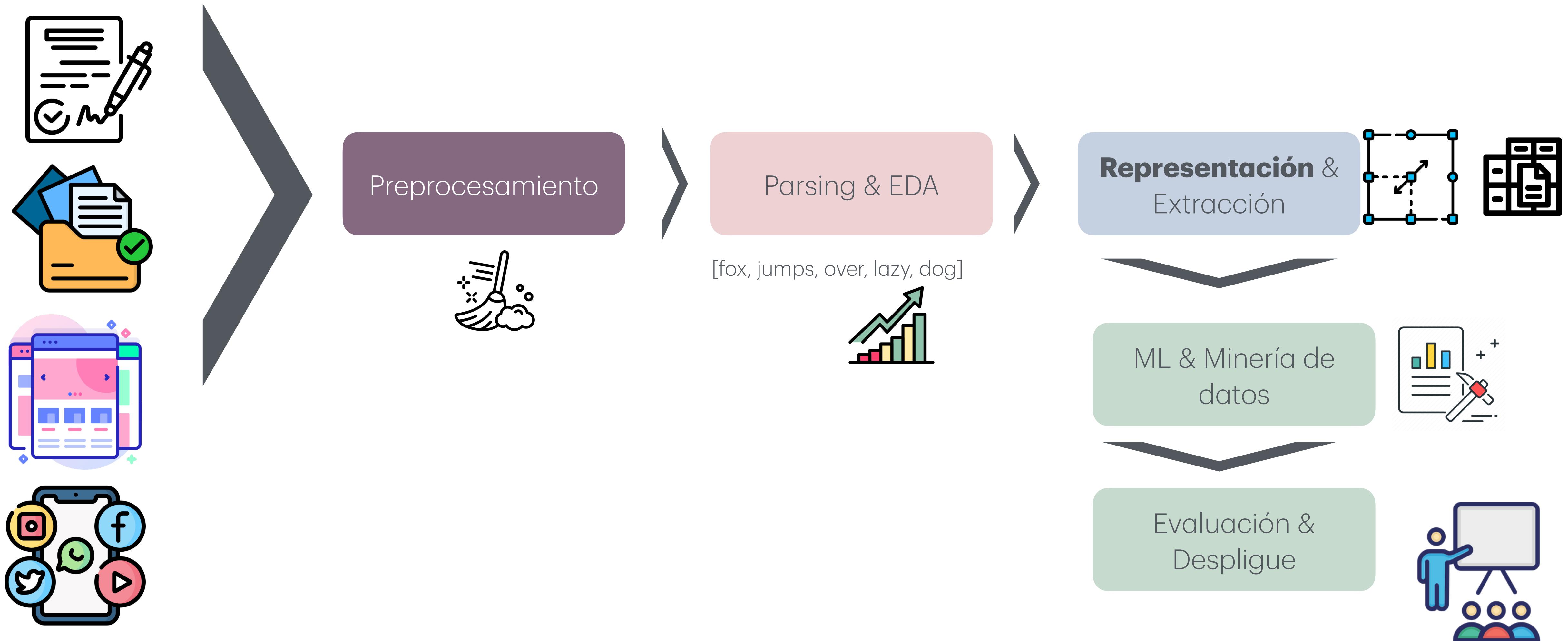
Análisis de sentimientos

Taller: Introducción al Procesamiento de Lenguaje Natural

2. Representación del texto

2. Representación del texto

Flujo de trabajo básico en NLP



2. Representación del texto

Machine Learning

“... es una subárea de la inteligencia artificial relacionado con algoritmos que aprenden por sí mismo a partir del conocimiento contenido en los datos”

— **Raschka, S.**

“... significa extraer conocimiento de un conjunto de datos”.

— **Müller, A. C.**

2. Representación del texto

Machine Learning

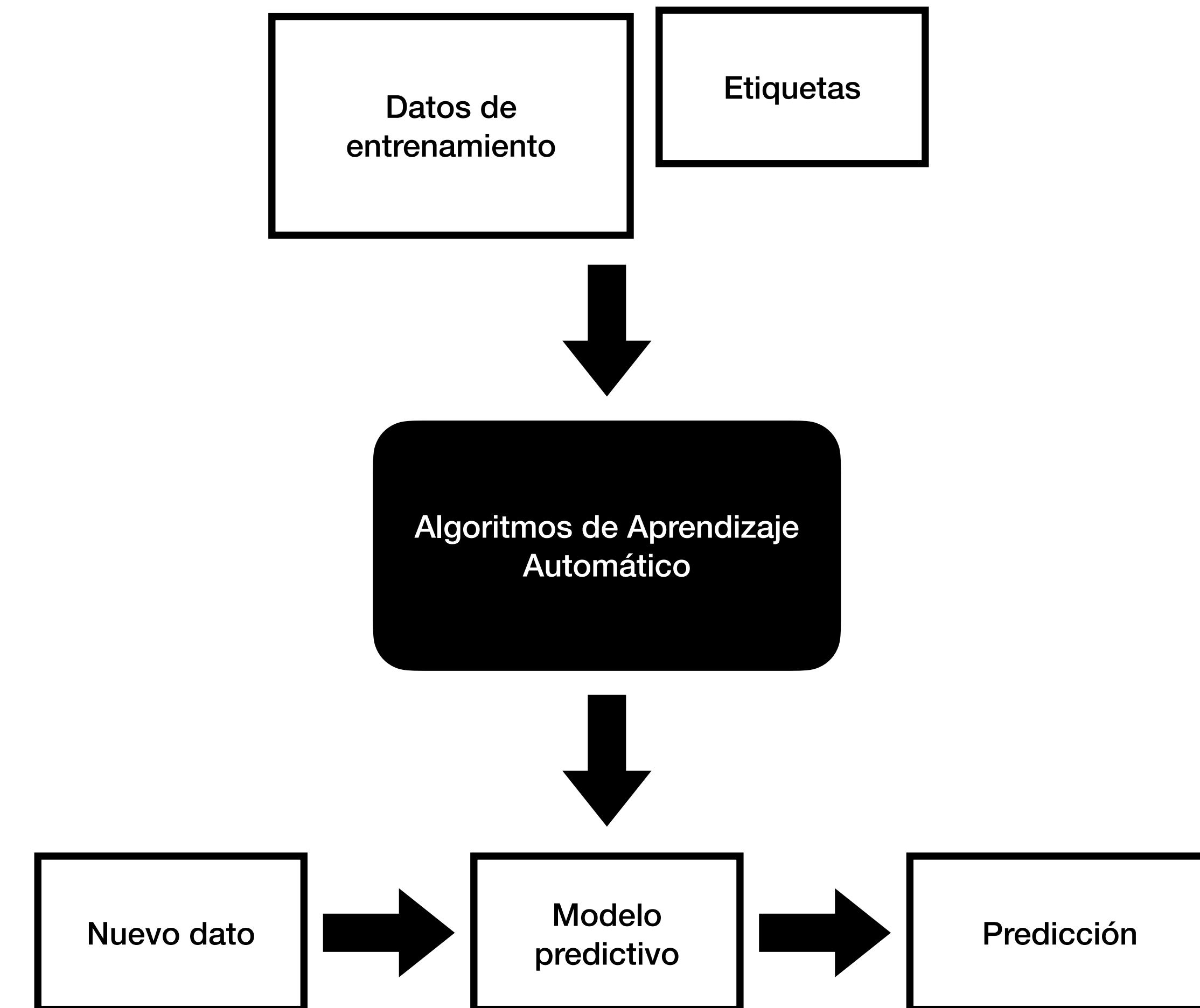
Los tres tipos de machine learning:

- Aprendizaje supervisado
 - Clasificación
 - Regresión
- Aprendizaje no supervisado
 - Agrupamiento
 - Reducción de dimensiones
- Aprendizaje de reforzamiento

2. Representación del texto

Machine Learning

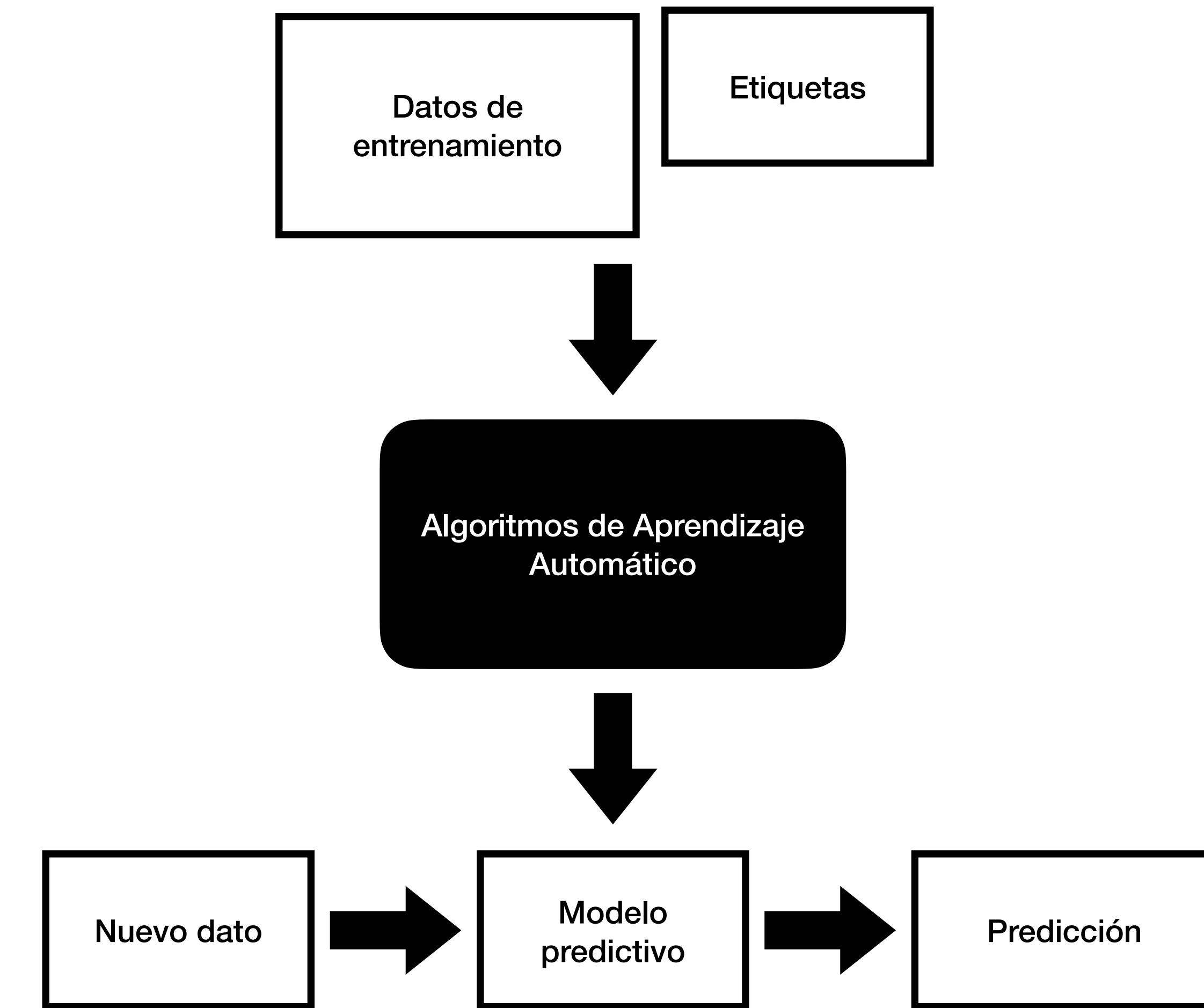
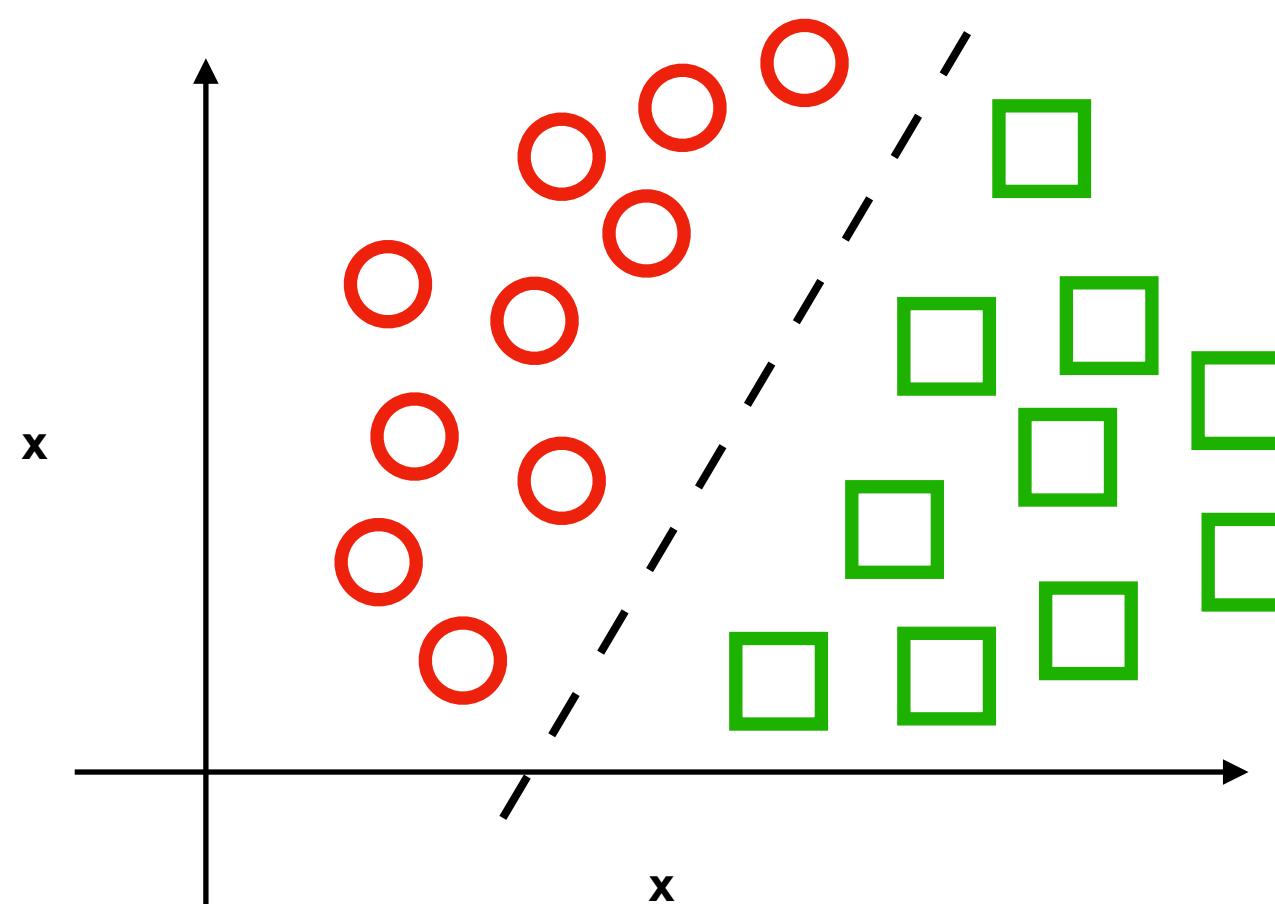
Aprendizaje supervisado



2. Representación del texto

Machine Learning

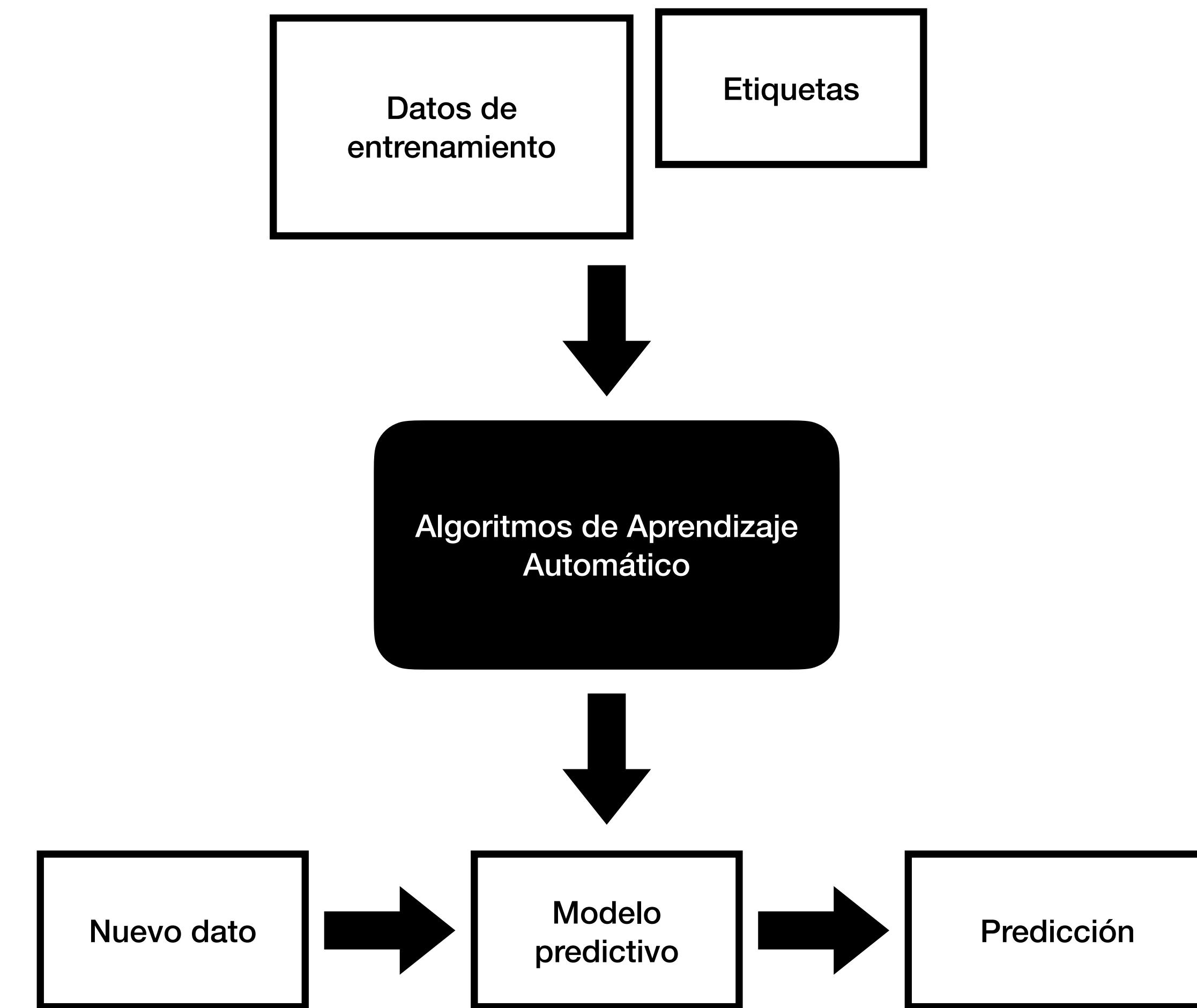
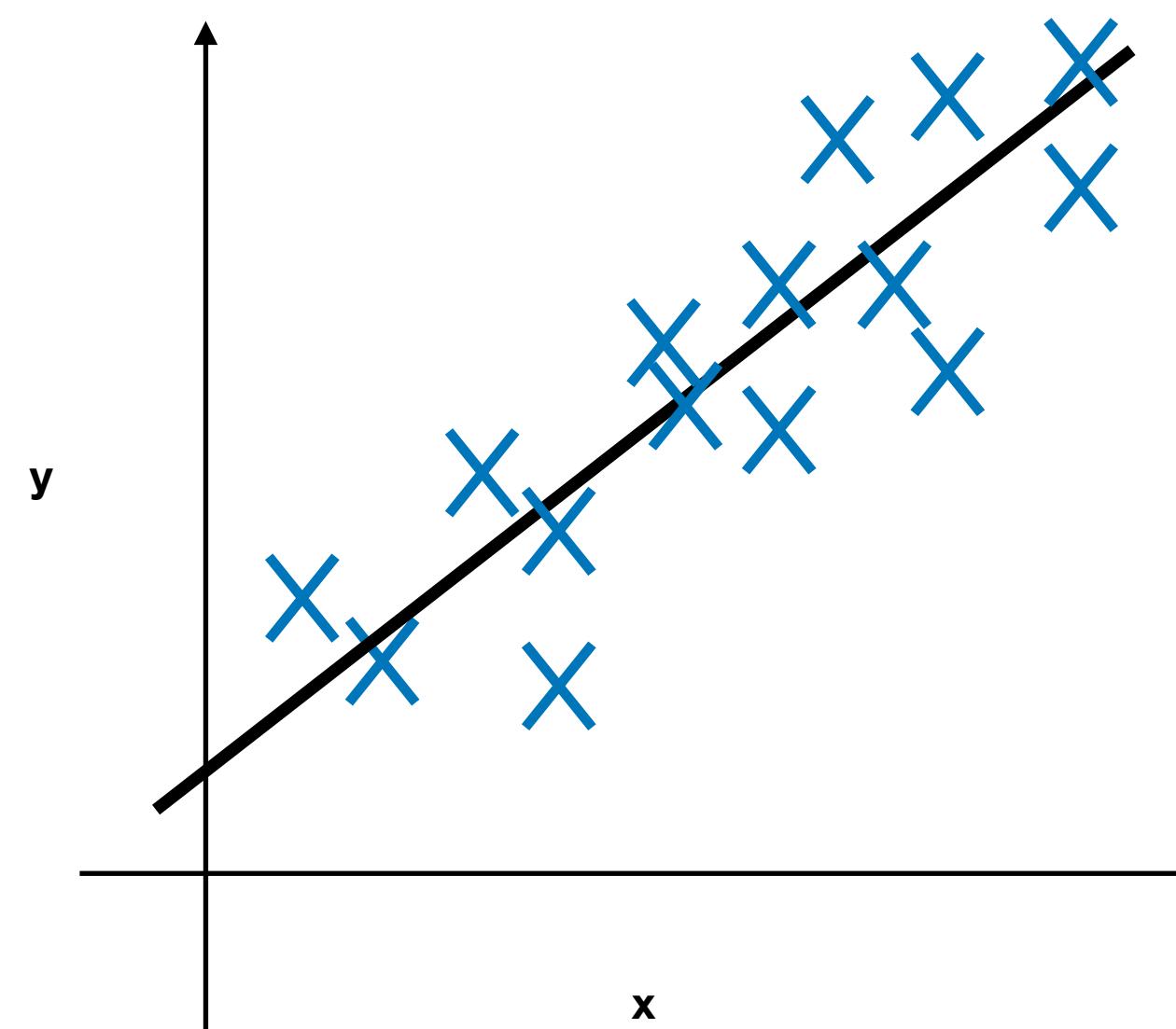
Aprendizaje supervisado



2. Representación del texto

Machine Learning

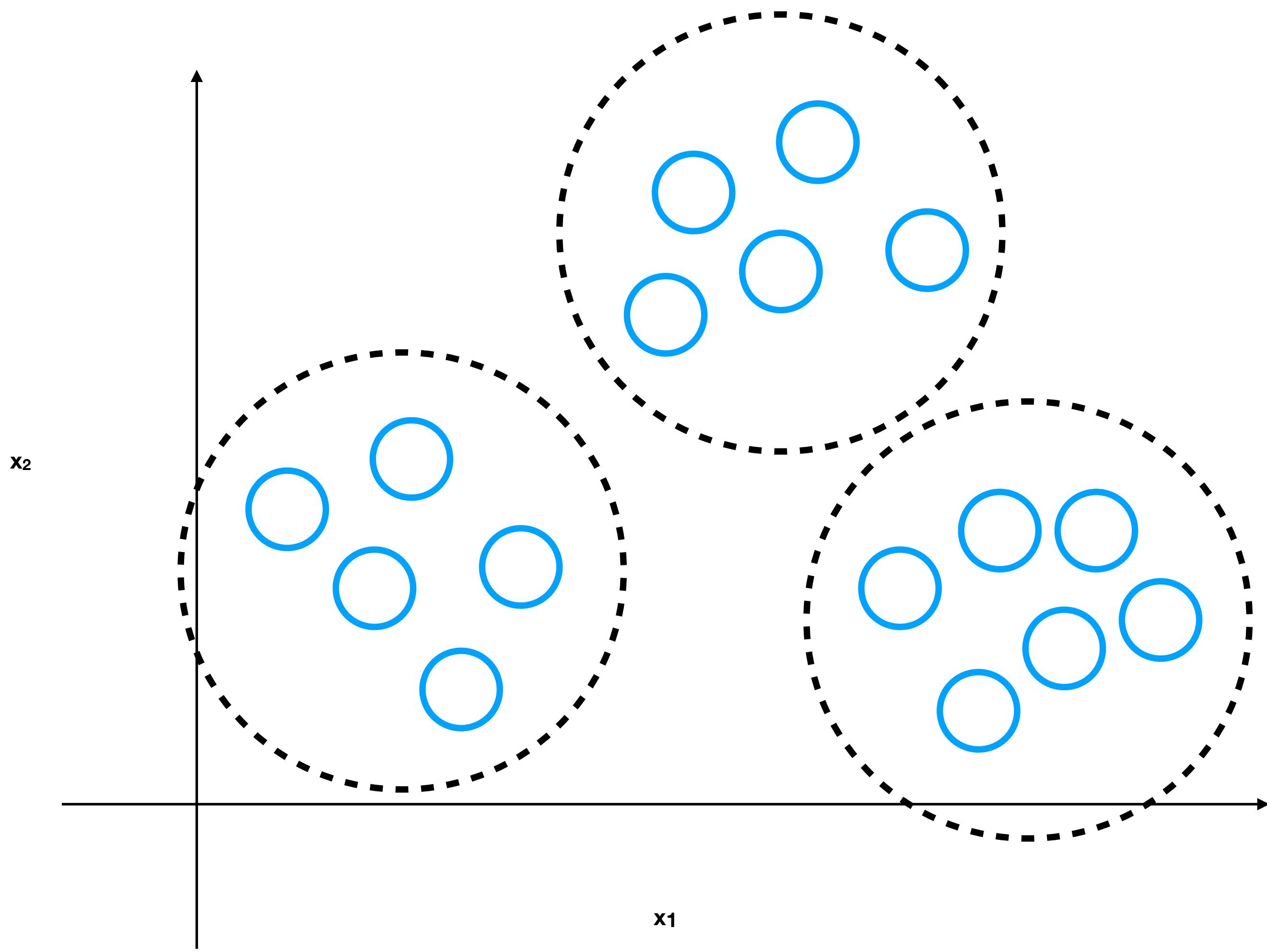
Aprendizaje supervisado



2. Representación del texto

Machine Learning

Aprendizaje no supervisado



2. Representación del texto

Machine Learning

Bases de datos

| Fisher's Iris Data | | | | |
|--------------------|-----------------|-----------------|-----------------|------------------|
| Largo de sépalo | Ancho de sépalo | Largo de pétalo | Ancho de pétalo | Especies |
| 5.1 | 3.5 | 1.4 | 0.2 | <i>I. setosa</i> |
| 4.9 | 3.0 | 1.4 | 0.2 | <i>I. setosa</i> |
| 4.7 | 3.2 | 1.3 | 0.2 | <i>I. setosa</i> |
| 4.6 | 3.1 | 1.5 | 0.2 | <i>I. setosa</i> |
| 5.0 | 3.6 | 1.4 | 0.2 | <i>I. setosa</i> |
| 5.4 | 3.9 | 1.7 | 0.4 | <i>I. setosa</i> |
| 4.6 | 3.4 | 1.4 | 0.3 | <i>I. setosa</i> |
| -- | -- | -- | -- | -- |



2. Representación del texto

Machine Learning

Bases de datos

Terminología y notaciones

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & x_4^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_4^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{(150)} & x_2^{(150)} & x_3^{(150)} & x_4^{(150)} \end{bmatrix}$$

Un ejemplo de una flor

$$x^{(i)} = [x_1^{(i)} \ x_2^{(i)} \ x_3^{(i)} \ x_4^{(i)}]$$

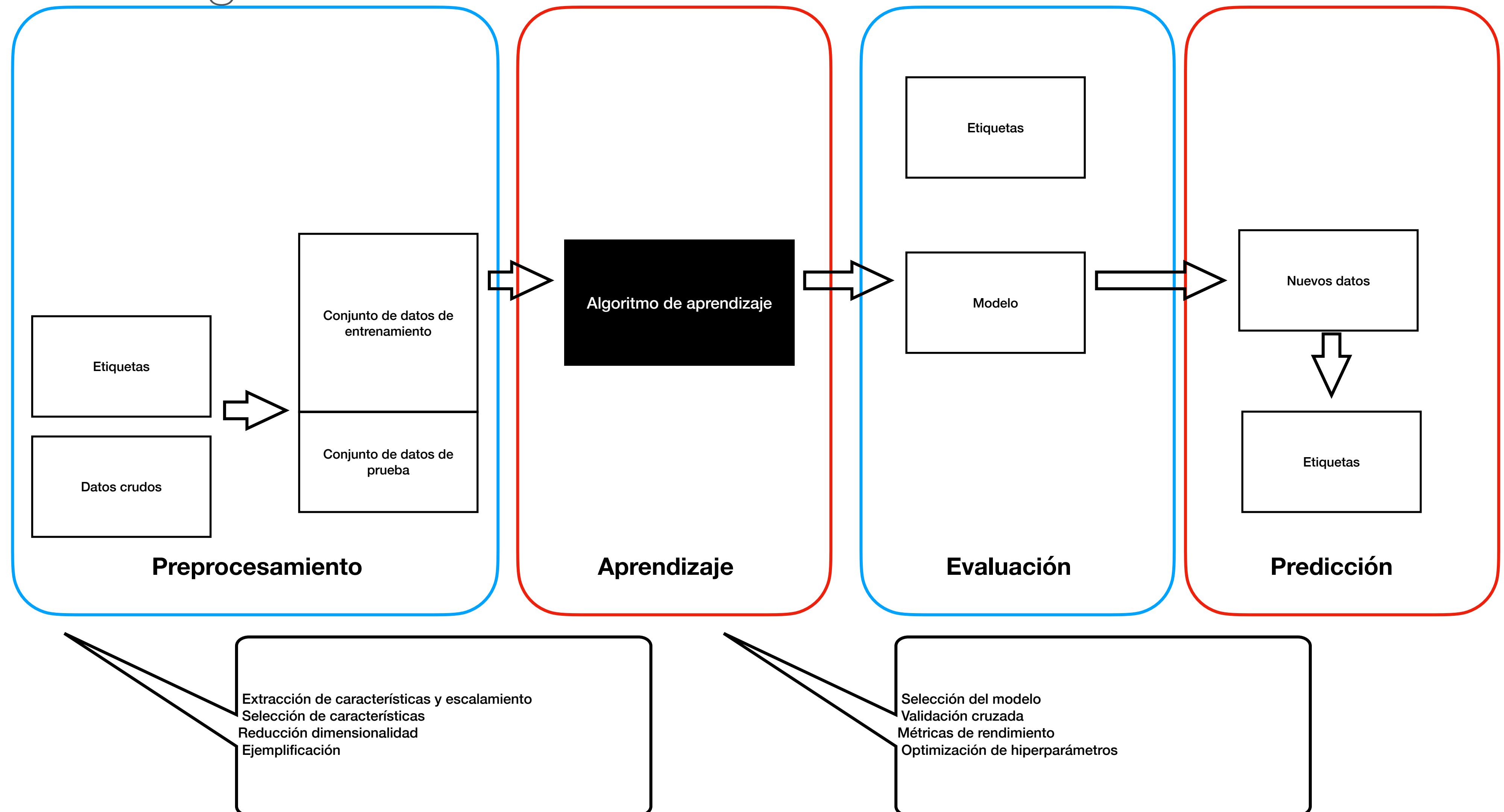
Variables objetivo o *targets*

$$y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(150)} \end{bmatrix}; y \in \{\text{Setosa}, \text{Versicolor}, \text{Virginica}\}$$



2. Representación del texto

Machine Learning



Taller: Introducción al Procesamiento de Lenguaje Natural

3. Aplicaciones básicas de NLP

Taller: Introducción al Procesamiento de Lenguaje Natural

4. Reflexiones