

OpenCourseWare
**Procesamiento de Lenguaje Natural con
Aprendizaje Profundo,**
Máster en Ciencia y Tecnología Informática

**Tema 1.1 Introducción: Procesamiento de
Lenguaje Natural y Aprendizaje Profundo**

Objetivos

- Presentar el campo del Procesamiento de Lenguaje Natural y conocer sus principales aplicaciones.
- Presentar los conceptos de aprendizaje automático y aprendizaje profundo.
- Comprender la diferencia entre los enfoques tradicionales de aprendizaje automático (algoritmos probabilísticos y estadísticos) y los enfoques de aprendizaje profundo.
- Conocer el potencial del uso del aprendizaje profundo en PLN.

Índice

- ¿Qué es Procesamiento de Lenguaje Natural (PLN)?.
- Principales aplicaciones de PLN.
- ¿Qué es aprendizaje profundo?

Procesamiento de Lenguaje Natural

- Técnicas computacionales para analizar y representar el lenguaje humano de forma automática
- Campo multidisciplinar: informática, lingüística, matemáticas, psicología, etc.

¿Por qué PLN?



Fuente: Gerald, [Pixabay](#)

- Crecimiento exponencial de los datos
 - 2013, 3.5 ZB
 - 2022, 40 ZB
 - 2025, 180 ZB
- > 80% en formato no estructurado, principalmente texto.

*1 ZB = 1 trillion GB

Índice

- ¿Qué es Procesamiento de Lenguaje Natural (PLN)?.
- **Principales aplicaciones de PLN.**
- ¿Qué es aprendizaje profundo?

Análisis morfosintáctico (PoS tagging) y de dependencias

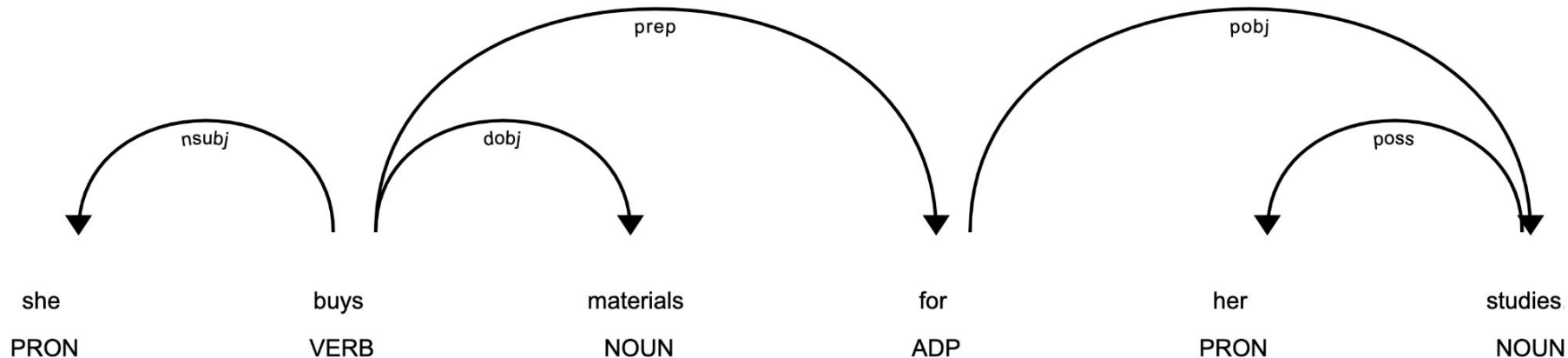
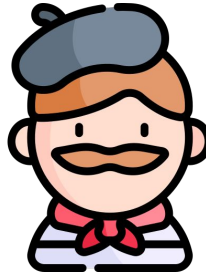


Imagen generada con la librería [Spacy](https://spacy.io/usage/visualizers) (<https://spacy.io/usage/visualizers>)

Detección del idioma



Fuente: Freepik [flaticon](#)



Fuente: Freepik [flaticon](#)



Fuente: Freepik [flaticon](#)



Fuente: Freepik [flaticon](#)



Fuente: Freepik [flaticon](#)

Traducción automática

Texto original:

Mein Name ist Isabel, ich komme aus Spanien. Ich bin 49 Jahre alt und habe zwei Töchter.



Texto traducido:

My name is Isabel, I'm from Spain. I am 49 years old and I have two daughters.

Recuperación de Información



Fuente: firmbee, [Pixabay](#)

Extracción de Información

Isabel Segura-Bedmar: HULAT at SemEval-2023 Task 10: Data Augmentation for Pre-trained Transformers Applied to the Detection of Sexism in Social Media. SemEval@ACL 2023: 184-192. 13-14 July 2023 Toronto.

Autor: Isabel Segura-Bedmar.

Título: HULAT at SemEval-2023 Task 10: Data Augmentation for Pre-trained Transformers Applied to the Detection of Sexism in Social Media.

Conferencia: SemEval@ACL 2023

Páginas: 184-192

Fechas: 13-14 July 2023

Lugar: Toronto

Extracción de Información. Tareas:

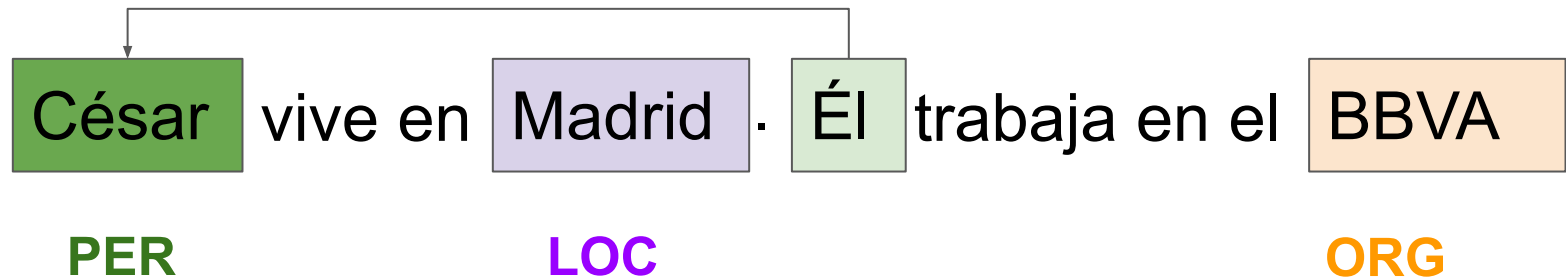
Reconocimiento de Entidades Nombradas. (Named Entity Recognition (NER))

César vive en Madrid . Él trabaja en el BBVA

PER LOC ORG

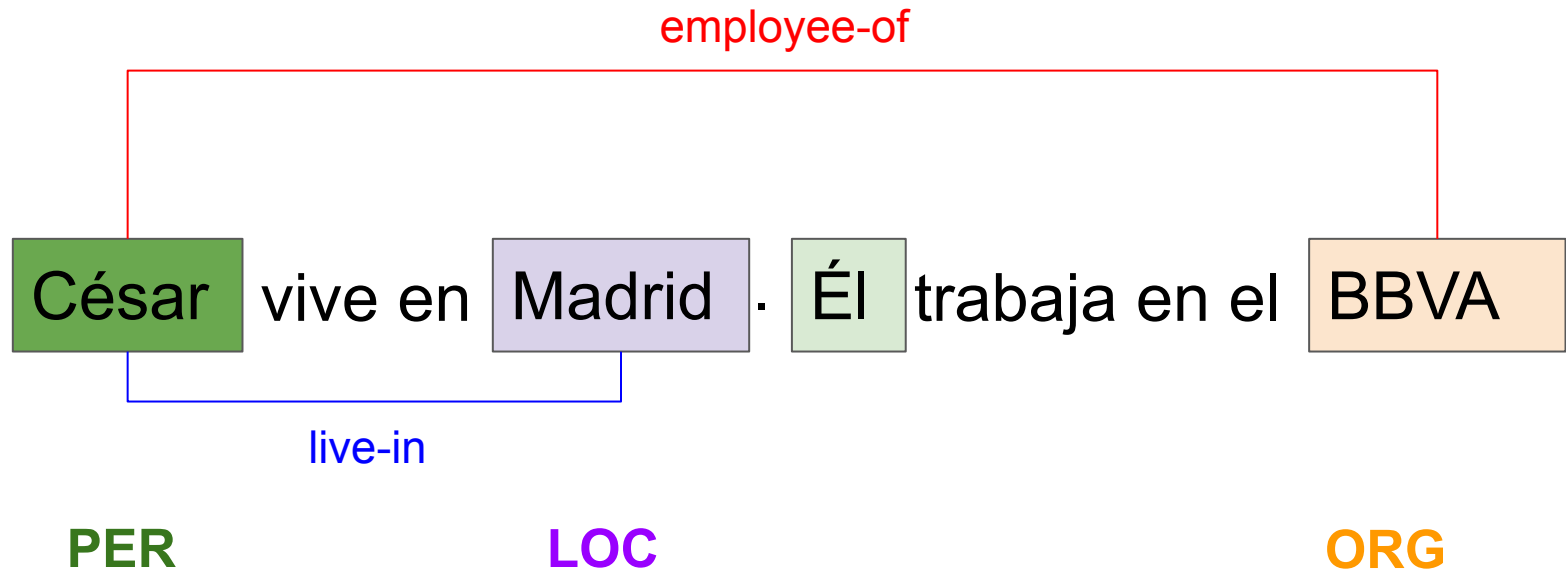
Extracción de Información. Tareas:

Resolución de la Correferencia



Extracción de Información. Tareas:

Extracción de relaciones



Sistemas de Búsqueda de Respuestas (Question Answering)



Fuente: Matheus Bertelli, [pexel](#)

Clasificación de Textos

- Múltiples aplicaciones. Algunos ejemplos son:
 - Detección de correo spam, noticias falsas, mensajes de odios en redes sociales, etc.
 - Clasificación de noticias.
 - Análisis de sentimiento.
 - Clasificación de notas clínicas.
 - Selección automática de cohortes de pacientes.

Generación de Resúmenes y Simplificación de textos

Texto original:

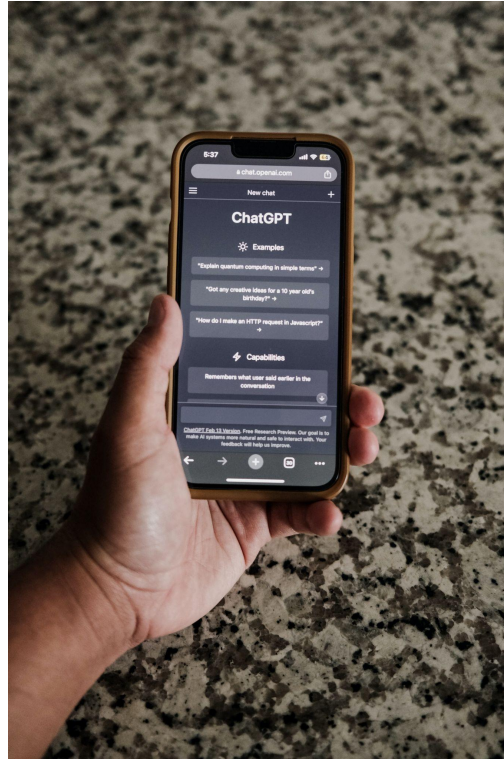
Juan tomó un taxi para ir a una reunión en el ministerio. Durante dicha reunión, Juan tuvo un ataque al corazón y fue trasladado al hospital.



Resumen:

Juan fue hospitalizado después de asistir a una reunión en el ministerio.

Asistentes Conversacionales (Chatbots)



Fuente: Airam Dato-on: [pexels](https://pexels.com)

Índice

- ¿Qué es Procesamiento de Lenguaje Natural (PLN)?.
- Principales aplicaciones de PLN.
- **¿Qué es aprendizaje profundo?**

Inteligencia Artificial:

programas con la capacidad de aprender y razonar como seres humanos.

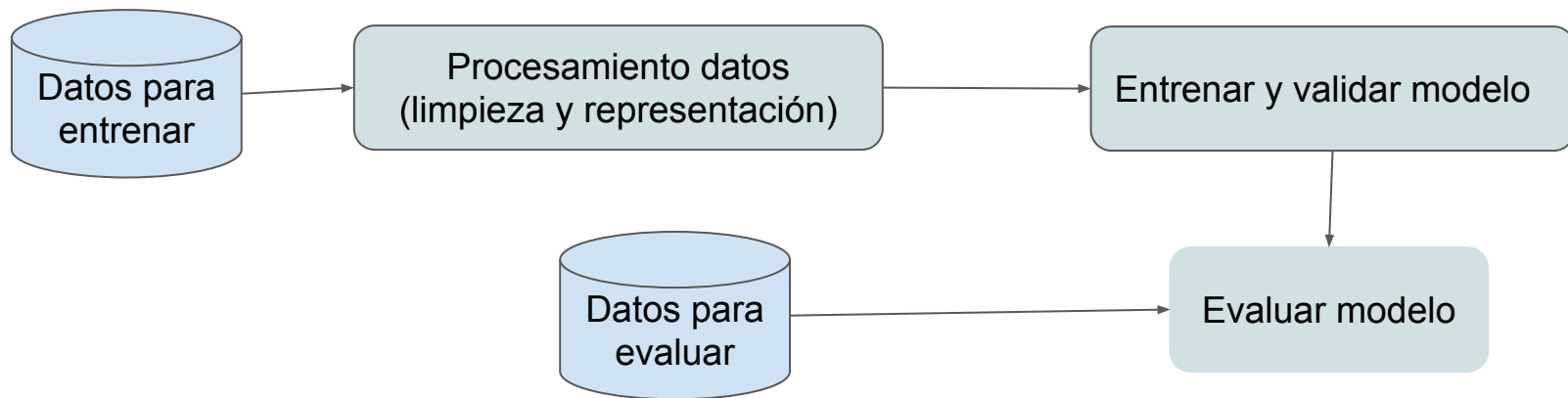
Aprendizaje automático:

algoritmos con la capacidad de aprender sin ser programados explícitamente

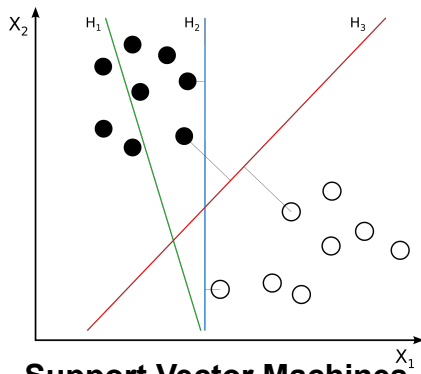
Aprendizaje profundo:

algoritmos de aprendizaje automático basados en redes neuronales con varias capas oculta

Aprendizaje automático (workflow)

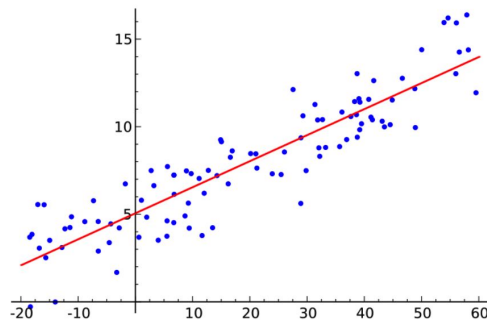


Aprendizaje automático



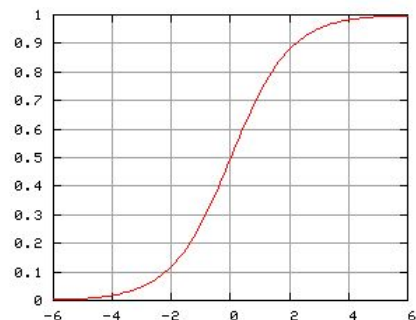
Support Vector Machines

Fuente: ZackWeinberg, [Wikimedia](#)



Regresión Lineal

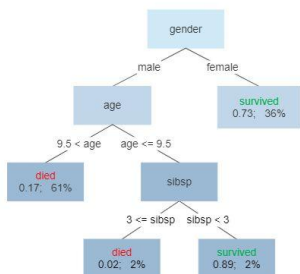
Fuente: Sewaqu, [Wikimedia](#)



Logistic regression

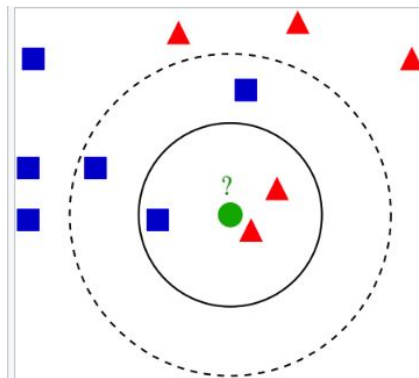
Fuente: [Wikimedia](#)

Survival of passengers on the Titanic



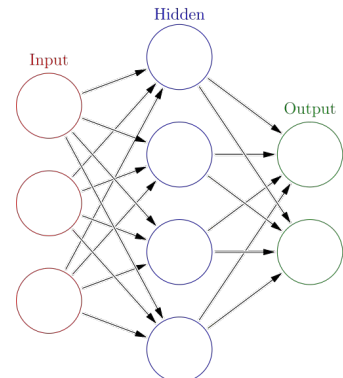
Árboles de Decision

Fuente: Gilgoldm, [Wikimedia](#)



k-nearest neighbors

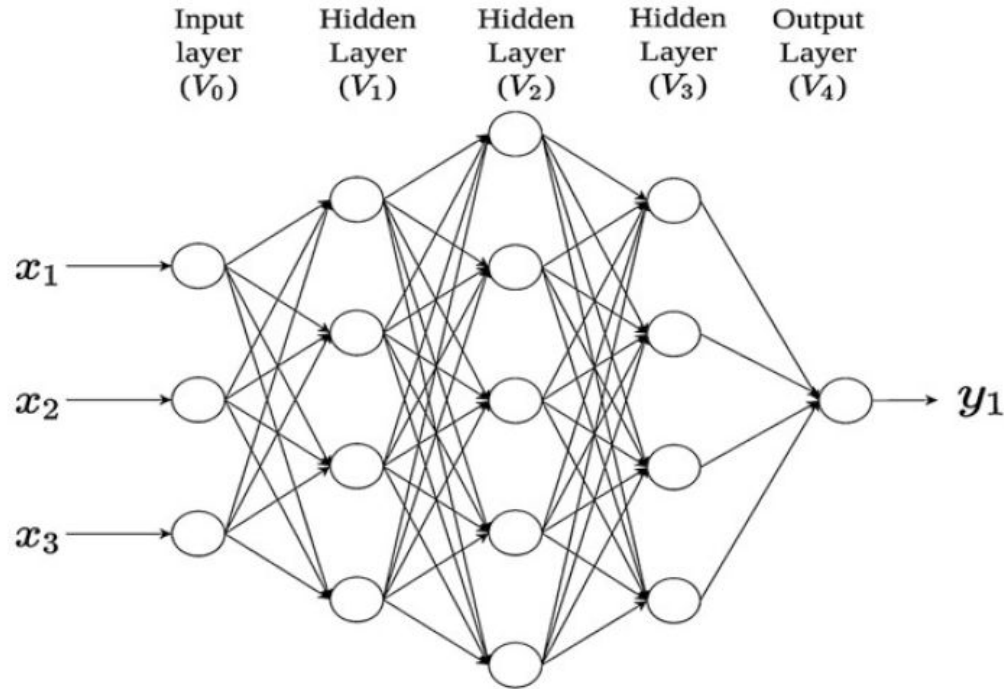
Fuente: Antti Ajanki AnAj, [Wikimedia](#)



Redes neuronales

Fuente: Glosser.ca, [Wikimedia](#)

Aprendizaje profundo: redes neuronales profundas



Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Fundamentals of artificial neural networks and deep learning. In *Multivariate statistical machine learning methods for genomic prediction* (pp. 379-425). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_10. CC BY 4.0.

Aprendizaje automático vs Aprendizaje profundo

- En los algoritmos de aprendizaje automático tradicionales, es necesario representar las instancias utilizando un conjunto de características.
- La definición de este conjunto de características (**feature extraction**) implica el conocimiento de expertos (tiempo y dinero).
- Los modelos de **aprendizaje profundo** son capaces de **aprender** de forma **automática** las mejores **características** para representar las instancias para una determinada tarea.
- El **aprendizaje profundo** ha supuesto una **revolución para PLN**, consiguiendo batir récords de resultados en muchas de sus aplicaciones.

Resumen

- **PLN**: entender y generar el lenguaje humano de forma automática.
- **Aprendizaje profundo** son redes neuronales con múltiples capas ocultas.
- La definición de características es una tarea crucial en los enfoques tradicionales de aprendizaje automático.
- **El aprendizaje profundo evita** la tarea de **definición de características** porque la red aprende la mejor representación de los datos.
- El aprendizaje profundo ha supuesto una **revolución en PLN**, obteniendo los mejores resultados en muchas tareas.

OpenCourseWare
Procesamiento de Lenguaje Natural con
Aprendizaje Profundo,

Gracias!!!

<https://github.com/isegura>