

## Lección 1: Sistemas de clasificación basados en textos.



## Desarrollo de la sesión

**En procesamiento del lenguaje natural (NLP)**, los sistemas de clasificación basados en textos emergen como una herramienta esencial para estructurar y comprender la ingente cantidad de datos textuales disponibles en el mundo digital. En esencia, estos sistemas permiten categorizar automáticamente documentos o textos en diferentes clases o categorías, lo que facilita enormemente la organización y el análisis de información en diversas aplicaciones.

La relevancia de los sistemas de clasificación basados en textos radica en su capacidad para automatizar un proceso que de otra manera sería laborioso y propenso a errores humanos. Imaginemos la tarea de etiquetar manualmente miles de correos electrónicos como "spam" o "no spam", o de clasificar documentos empresariales en distintas categorías según su contenido. Esta labor sería no solo tediosa, sino también poco práctica en un entorno donde la velocidad y la eficiencia son críticas.

En este sentido, los sistemas de clasificación basados en textos se convierten en aliados indispensables en numerosos contextos. Desde el ámbito empresarial, donde facilitan la gestión y organización de documentos, hasta la investigación científica, donde ayudan a identificar estudios relevantes en vastas bases de datos académicas, estos sistemas desempeñan un papel fundamental en la optimización de procesos y la extracción de conocimientos significativos.

La seguridad de la información se convierte en una preocupación creciente. Los sistemas de clasificación basados en textos también tienen un rol crucial en este ámbito, contribuyendo a la detección de contenido malicioso como correos electrónicos de phishing, spam o incluso la identificación de amenazas cibernéticas más sofisticadas.