



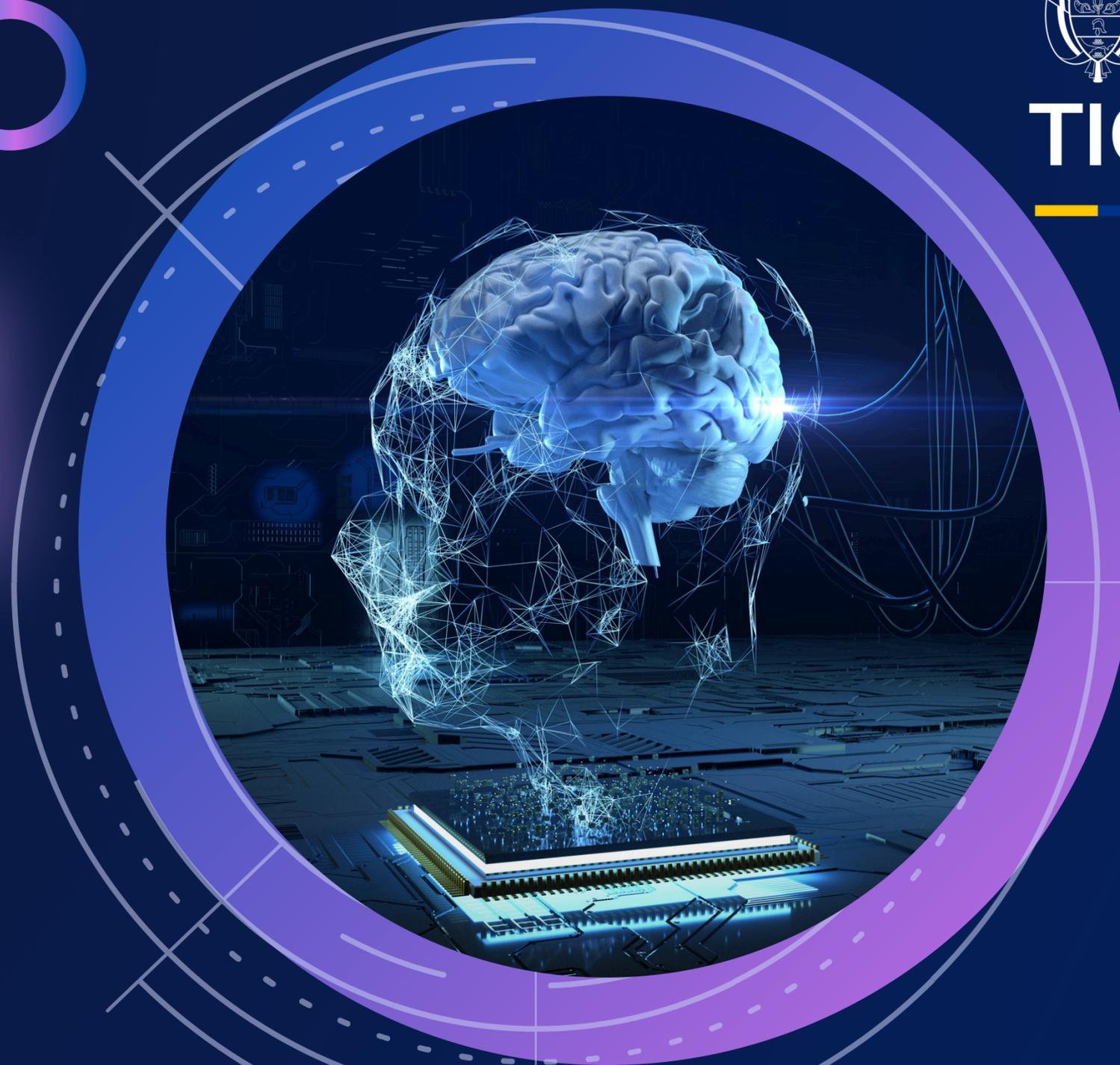
TIC



▶ TALENTO TECH

**BOOTCAMP
ANÁLISIS DE DATOS**

INTEGRADOR - Módulo 3



Contextualización de mis aprendizajes

El análisis de texto abarca diversos aspectos esenciales del procesamiento del lenguaje natural, incluyendo sistemas de clasificación basados en textos, uso de palabras separadas, metodologías de análisis de sentimientos en textos y análisis de sentimientos con bases de datos en Twitter. Estas áreas son fundamentales para comprender y extraer información significativa de grandes volúmenes de datos textuales.

Los sistemas de clasificación basados en textos son herramientas poderosas que permiten categorizar documentos o textos en diferentes clases o categorías. Estos sistemas utilizan técnicas como el aprendizaje automático para entrenar modelos que pueden asignar automáticamente etiquetas o clasificaciones a nuevos textos en función de características específicas. La eficacia de estos sistemas depende en gran medida de la calidad de los datos de entrenamiento y de la selección adecuada de características relevantes para la clasificación.



TIC



Contextualización de mis aprendizajes

El uso de palabras separadas es una técnica común en el análisis de texto que implica descomponer textos en palabras individuales. Esta técnica es fundamental para realizar análisis de frecuencia de palabras, identificar patrones lingüísticos y extraer características importantes para diferentes tareas, como la clasificación de documentos o el análisis de sentimientos. Al separar el texto en palabras, se facilita el procesamiento y la comprensión del contenido textual.

Las metodologías de análisis de sentimientos en textos se centran en identificar y clasificar las emociones expresadas en el texto. Estas metodologías pueden variar desde enfoques basados en reglas hasta modelos de aprendizaje automático más sofisticados. El análisis de sentimientos es útil en una variedad de aplicaciones, como la monitorización de la opinión pública en redes sociales, la evaluación de la satisfacción del cliente y la detección de tendencias en la opinión del usuario.



TIC





TIC



El análisis de sentimientos con bases de datos en Twitter es un área específica dentro del análisis de sentimientos que se enfoca en extraer y analizar opiniones y emociones expresadas en los tweets. Dada la naturaleza rápida y dinámica de Twitter, este tipo de análisis puede proporcionar información valiosa sobre eventos actuales, tendencias de opinión y la percepción del público sobre diversos temas. Las técnicas utilizadas en este contexto pueden incluir el procesamiento de lenguaje natural, el análisis de redes sociales y el aprendizaje automático para extraer y analizar datos de manera eficiente y precisa.





UNIDAD 3

Objetivo General: Al concluir esta lección, los estudiantes habrán adquirido las habilidades necesarias para:

- Comprender los fundamentos de los sistemas de clasificación basados en textos y su aplicación en la categorización automatizada de documentos o textos.
- Utilizar de manera efectiva técnicas de separación de palabras para analizar y procesar textos de manera más estructurada y comprensible.
- Aplicar diversas metodologías de análisis de sentimientos en textos para identificar y clasificar las emociones expresadas en diferentes contextos y tipos de texto.
- Realizar análisis de sentimientos con bases de datos en Twitter, extrayendo información valiosa sobre opiniones, tendencias y percepciones del público en esta plataforma de redes sociales.
- Integrar las habilidades adquiridas en estas áreas para abordar problemas complejos en el procesamiento del lenguaje natural y mejorar la comprensión y manipulación de datos textuales en diversos escenarios y aplicaciones.



Tiempo de ejecución: 2 horas



TIC



COMPETENCIAS A DESARROLLAR

- Comprensión de sistemas de clasificación basados en textos.
- Utilización efectiva de técnicas de separación de palabras

Adquirir conocimientos sobre cómo funcionan los sistemas de clasificación basados en textos y su importancia en la categorización automática de documentos o textos.

Aprender a aplicar técnicas de separación de palabras para analizar y procesar textos de manera estructurada y comprensible.





TIC



COMPETENCIAS A DESARROLLAR

- Aplicación de metodologías de análisis de sentimientos en textos.
- Realización de análisis de sentimientos con bases de datos en Twitter.
- Integración de habilidades para abordar problemas complejos en nlp.

Desarrollar habilidades en la aplicación de metodologías de análisis de sentimientos para identificar y clasificar emociones en diferentes contextos textuales.

Practicar la extracción y análisis de datos de opiniones, tendencias y percepciones del público en Twitter mediante el análisis de sentimientos.

Combinar las competencias adquiridas para abordar problemas complejos en el procesamiento del lenguaje natural y mejorar la comprensión y manipulación de datos textuales en diversas aplicaciones.



Activación de saberes previos



TIC

PLANTEAMIENTO DE LA SESIÓN

Los sistemas de clasificación basados en textos son herramientas fundamentales en el campo del procesamiento del lenguaje natural (PLN) que permiten categorizar automáticamente documentos o textos en diferentes clases o categorías. Estos sistemas son ampliamente utilizados en una variedad de aplicaciones, como la clasificación de correos electrónicos, la detección de spam, la organización de documentos y la categorización de noticias, entre otros. Conceptos básicos de clasificación de textos: Se introducirán los conceptos básicos de la clasificación de textos, incluyendo la definición de clases o categorías, características del texto y el proceso de entrenamiento y evaluación de un modelo de clasificación.

MATERIALES

PC con conexión a internet.



PLANTEAMIENTO DE LA SESIÓN	MATERIALES
<p>Representación de documentos: Se explorarán diferentes técnicas para representar documentos de texto de manera que sean adecuados para su procesamiento por parte de algoritmos de aprendizaje automático.</p> <p>Algoritmos de clasificación: Se presentarán varios algoritmos de clasificación utilizados comúnmente en sistemas de clasificación basados en textos, como Naive Bayes, Support Vector Machines (SVM), y algoritmos de aprendizaje profundo como las redes neuronales convolucionales (CNN) o las redes neuronales recurrentes (RNN).</p> <p>Entrenamiento y evaluación de modelos: se aprenderá sobre el proceso de entrenamiento de un modelo de clasificación utilizando conjuntos de datos etiquetados y técnicas de validación cruzada para evaluar el rendimiento del modelo.</p>	