Tarea #3  
Análisis de sentimientos con NLTK en Google Colab

línea corta

**PONDERACIÓN: 2 puntos**

**Horas Aproximadas: 5 horas**

Universidad San Carlos de Guatemala

Facultad de ingeniería.

Ingeniería en ciencias y sistemas

# Introducción

EL uso de Python y sus librerías para el análisis facilita el desarrollo de este, Python ha evolucionado a tal punto que es uno de los mejores lenguajes para este tipo de actividades, el estudiante debe utilizar herramientas las cuales le permitan poder analizar los sentimientos de texto de un archivo, agregado de crear un documento utilizando Google Colab para interactuar con herramientas más modernas.

# Competencia(s)

# Construye modelos predictivos y análisis de sentimientos usando bibliotecas Python como Scikit-learn y NLTK siguiendo la metodología CRISP-DM para resolver problemas reales a partir de datos estructurados/textuales

# Contenido

# El estudiante utilizará Google Colab para procesar texto y aplicar análisis de sentimientos utilizando NLTK. Deberá:

# 1. Cargar un conjunto de opiniones o reseñas (puede ser de productos, servicios, películas).

# 2. Realizar limpieza del texto (tokenización, eliminación de stopwords, lematización).

# 3. Implementar un modelo de análisis de sentimientos simple (clasificar entre reseñas positivas y negativas usando conteo de palabras o un modelo Naive Bayes).

# 4. Presentar resultados básicos (porcentaje de reseñas positivas/negativas) y mostrar ejemplos.

# 5. Entregar el notebook en Colab con su ejecución completa y breve análisis escrito.

# Recursos Adicionales

Google Colab: https://colab.google/

NLTK API: https://www.nltk.org/

# Forma de entrega

Enlace del repositorio del curso nombrado “SS2\_2S2025\_#carné” en la carpeta Tarea\_3 con un archivo README con toda la documentación requerida y los archivos necesarios (código, imágenes, JSON, XML) dependiendo del criterio del estudiante. (Obligatorio enlace del notebook)

# Fecha de entrega

Fecha límite para realizar la entrega de la tarea

# Rúbrica de Calificación

|  |  |  |
| --- | --- | --- |
| **Criterio** | **Descripción** | **Puntos Máximos** |
| **Carga del dataset** | Se carga correctamente un conjunto de opiniones/reseñas reales (propias o de una fuente válida como Kaggle, IMDb, Amazon, etc.). | 10 pts |
| **Limpieza de texto** | Se aplican correctamente al menos tres técnicas de preprocesamiento: tokenización, eliminación de stopwords, lematización. | 20 pts |
| **Modelo de análisis de sentimientos** | Implementa correctamente un clasificador simple (por conteo de palabras o Naive Bayes). Se debe entrenar y usar para clasificar. | 25 pts |
| **Presentación de resultados** | Muestra métricas básicas (porcentaje de positivos/negativos) y ejemplos de predicciones (casos de reseñas reales con clasificación). | 15 pts |
| **Uso de Google Colab y ejecución clara** | El notebook funciona, tiene celdas ejecutadas, explicaciones, código limpio y resultados visibles sin errores. | 15 pts |
| **Informe/README en repositorio** | Contiene una breve explicación del proceso, herramientas utilizadas y análisis general. Indica el enlace funcional del notebook. | 10 pts |
| **Formato de entrega correcto** | El repositorio está correctamente nombrado, organizado y ubicado en la carpeta correspondiente según las instrucciones. | 5 pts |

# Valores

|  |  |  |
| --- | --- | --- |
| **Valor** | **Aplicación en esta tarea** | **Consecuencia si no se cumple** |
| **Organización** | El notebook debe estar ordenado, con secciones claras, y el repositorio debe tener todos los archivos necesarios en su lugar. | ❌ Hasta **-10 puntos** si el notebook está desordenado o el repositorio incompleto. |
| **Responsabilidad** | Se espera la ejecución completa y funcional del notebook, y que se entregue dentro del tiempo establecido. | ❌ **0 puntos** si el notebook no corre o está completamente vacío. |
| **Disciplina** | El estudiante debe mostrar que aplicó paso a paso el flujo de CRISP-DM, desde la preparación del texto hasta el análisis. | ❌ Hasta **-15 puntos** si el flujo está incompleto o no justificado. |
| **Pensamiento lógico** | El modelo y el análisis deben tener sentido lógico, relacionando bien los pasos de limpieza, entrenamiento y resultados obtenidos. | ❌ Hasta **-10 puntos** si el análisis es incoherente o no muestra comprensión. |