



Universidad del Valle
Maestría en Analítica e Inteligencia de Negocios
Analítica de Datos en Salud

TAREA No. 1: Extracción de Conceptos Médicos

Para desarrollar esta actividad usted debe crear una herramienta de extracción de información que permita procesar historias clínicas de cáncer de pulmón de manera automática. Debe utilizar los conceptos de **Fine-Tuning** y **Transfer-Learning**. Debe integrar un modelo pre-entrenado de lenguaje natural con un corpus anotado para extracción de entidades nombradas.

Entregable 1: Un script documentado en Python donde se muestre el proceso de afinamiento del corpus de cáncer de pulmón que se encuentra en el siguiente enlace.

<https://drive.google.com/drive/folders/1DBnlco-vBtEXktGQCR7QR4vwVv8JMw46>

Entregable 2: Un script documentado en Python donde se muestre el proceso de validación del modelo creado en el punto anterior. Este script debe permitir cargar el modelo y usarlo para extraer información de un conjunto de oraciones de entrada. Utilice como ejemplo 10 oraciones y analice los resultados que produce el modelo. El modelo se comporta correctamente?

Los dos entregables anteriores se deben presentar ante el profesor y los compañeros de clase. Se harán preguntas individuales a cada integrante del grupo (Máximo 4 integrantes)