Text Mining & Image Recognition
Proyecto Final

Instrucciones: A continuaciÃn verÃ¡ dos ejercicios que debe completar para poder entregar su proyecto final. PodrÃ¡ realizar su cÃdigo en un Notebook con los dos ejercicios. DeberÃ¡ entregar sus ejercicios por medio de github.

Problema 1 - Word Cloud:
Descargue el Dataset (de click aquÃ para descargar) el cual contiene aproximadamente 800,000 tweets de diversos temas.

Usando CoLab y expresiones regulares. Determine los 3 usuarios mÃ¡s populares dentro del dataset. Luego arme un corpus el cual contenga los siguientes elementos por cada usuario seleccionado:

- Content: Tweet.
- Metadata: ID, Timestamp, Length (este valor hay que calcularlo).

Posterior a tener sus 3 corpus creados, responda: Â¿RazÃn por la que citan a ese usuario? para esto es necesario que extraiga el contexto de cada tweet y verifique cuales son las palabras que m ?as rodean al nombre de usuario. Para extraer un contexto valido y debido a la naturaleza del tipo de datos que estÃ¡n disponibles en nuestro dataset le recomendamos seguir los siguientes pasos:

1. Remover stopwords.
2. Realizar stemming y lemmatizaciÃn.
3. Mostrar un wordcloud con el top 10 para cada usuario.

Problema 2 - Fruits and Vegetables Recognizer:

Desarrollar un modelo de clasificación de imágenes basado en redes neuronales convolucionales (CNNs) capaz de distinguir entre diferentes tipos de frutas y verduras, utilizando un conjunto de datos inicial centrado en imágenes. El modelo deberá ser robusto ante variaciones en tamaño, rotación y condiciones de iluminación de las imágenes.

Parte #1: Dataset:

Deberá descargar el dataset de frutas disponible en este link de (Kaggle). Luego deberá seleccionar al menos tres frutas y tres vegetales para desarrollar su proyecto.

- Data Augmentation y Preprocesamiento: Aplicar técnicas de aumento de datos para generar nuevas imágenes a partir de las existentes (rotación, escalado, flip horizontal, ruido entre otras.) adicionalmente deberá considerar si resolverá el problema utilizando escala de grises o imágenes a color. Por ultimo deberá redimensionar las imágenes a un tamaño estándar para la entrada de la red neuronal convolucional así como normalizar los valores a valores entre 0 y 1, todo esto es posible realizarlo por medio de la función ImageDataGenerator y la función flow_from_directory de Keras.

Parte #2: Diseño y Entrenamiento de la Red Neuronal Convolucional:

- Arquitectura: Deberá considerar el número de capas convolucionales y de pooling, tipo de pooling, el tamaño de los filtros (kernels), la función de activación y el número de clases a clasificar, el numero de capas y neuronas artificiales, los parámetros de optimización y entrenamiento y demás parámetros para implementar su CNN. Deberá probar al menos 3 arquietcutas distintas para resolver el problema y validar cual de las 3 funciona mejor. y deberá contestar la pregunta Por qué cree que la CNN que seleccionó funciona mejor que otras?
- Entrenamiento: Deberá dividir el conjunto de datos en conjuntos de entrenamiento, validación y prueba para determinar la eficiencia de su algoritmo.