

Jorge Armando Castro Escudero

Cc. 98708993

Entrega 1

- **Contexto de aplicación**

En los últimos años se han reportado estudios sobre estabilización de suelos viales apoyados en la minería de datos y programas estadísticos para determinar relaciones estabilizantes/suelo que permitan obtener los mejores desempeños en el material a utilizarse en una estructura de pavimento. Algunos investigadores del área de suelos proponen minimizar el número de ensayos de laboratorio y crear estrategias de ubicación de muestreo mediante el desarrollo espacial de mapas de la resistencia a la compresión no confinada a diferentes niveles de tratamiento para cal y cemento, utilizando técnicas de aprendizaje automático y el uso de una base de datos compilada a partir de varios estudios sobre estabilización de suelos en los Estados Unidos.

Para efectos del curso se seleccionó un challenge de www.kaggle.com llamado "soil types", para emplear las redes convulsionales e imágenes de diferentes tipos de suelos a modo de entrenamiento, en mi trabajo de tesis se emplearán datos propios.

Context

The dataset consists of 5 varieties of soil images in 5 directories or folders. The dataset is a very small dataset meant for beginners to build ML models using this dataset currently. A small dataset helps learn without wasting hefty time in training the model for half an hour or more and expensive computation. The results won't be really great as the dataset is really low and thus overfitting and other issues.

- **Objetivo de machine learning**

En la clasificación de suelos a nivel de formación geológica existen varias categorías según su origen o color; para este ejercicio se consideran cinco categorías de la siguiente manera y con diferentes números de archivos.

Black soil: 37 imágenes, Cinder soil: 30 imágenes, Laterite soil: 30 imágenes, Peat soil: 30 imágenes, Yellow soil: 29 imágenes. Mediante el empleo de redes convulsionales, se desea predecir el tipo de suelo.

- **Dataset**

Para este ejercicio se emplean redes convulsionales, las imágenes fueron cargadas en el espacio de prueba del Colab, de acuerdo con una jerarquía de carpetas, luego se hace el cargue de las imágenes, se redimensionan y se etiquetan.

- Se importan las librerías de numpy y pandas para guardar en arreglos la información de cada tipo de imagen.
- Se utilizó tensorflow, el paquete de keras para redimensionar las imágenes
- Se reescalan las imágenes de 1/255 para manejar un RGB uniforme de las imágenes.
- Se subdividen los datos en: train y test
- Se hace el entrenamiento por medio de un batch de imágenes y se reagrupa de acuerdo con las categorías establecidas.

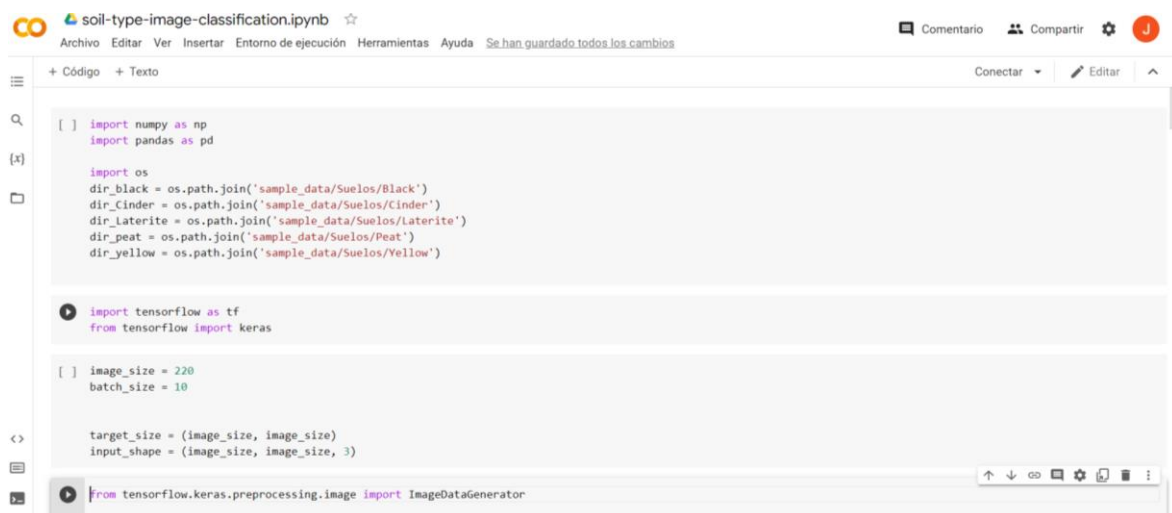
Black :0, Cinder: 1, Laterite:2, peat: 3, Yellow:4

- **Métricas de desempeño (de machine learning y negocio)**

- Se construye un modelo predictivo mediante la aplicación de una red neuronal convulsional de 128 neuronas y 5 salidas y mediante aplicación de las métricas "accuracy y epochs" se hará la evaluación del modelo para la predicción de datos futuros

- **Referencias y resultados previos**

A continuación, se adjunta pantallazo de los avances realizados con el database seleccionado.



```
[ ] import numpy as np
import pandas as pd

import os
dir_black = os.path.join('sample_data/Suelos/Black')
dir_cinder = os.path.join('sample_data/Suelos/Cinder')
dir_laterite = os.path.join('sample_data/Suelos/Laterite')
dir_peat = os.path.join('sample_data/Suelos/Peat')
dir_yellow = os.path.join('sample_data/Suelos/Yellow')

import tensorflow as tf
from tensorflow import keras

[ ] image_size = 220
batch_size = 10

target_size = (image_size, image_size)
input_shape = (image_size, image_size, 3)

from tensorflow.keras.preprocessing.image import ImageDataGenerator
```



```
from tensorflow.keras.preprocessing.image import ImageDataGenerator

train_datagen = ImageDataGenerator(rescale=1/255)

train_generator = train_datagen.flow_from_directory(
    'sample_data/Suelos/',
    target_size=(200, 200),
    batch_size = batch_size,
    classes = [ 'Black', 'Cinder', 'Laterite', 'Peat', 'Yellow'],
    class_mode='categorical')
```

Found 156 images belonging to 5 classes.

<https://www.kaggle.com/datasets/prasanshasatpathy/soil-types>, Información de la base de datos tomada de Kaggle: