

William Andrés Gómez Roa  
Bioingeniería - Ciencia de Datos

## Proyecto Final de IA: Análisis de Propiedades del Suelo y Absorbancia

### Video Demostrativo

<https://youtu.be/YIBE94VrF30>

¡Mira mi video explicativo para obtener una visión general de este proyecto!

### Introducción

En este proyecto final de IA, se realizó un análisis exhaustivo de las propiedades del suelo y la absorbancia en diferentes longitudes de onda para el dataset ‘NIRS-data.csv’. El objetivo principal fue comprender cómo las características del suelo, como pH, Al, Fe, Mg, Mn, Cu, B, entre otros minerales, influyen en la fertilización del suelo, guiada por el tipo de cultivo y la ubicación geográfica (Departamento).

### Metodología

El análisis se dividió en las siguientes etapas:

1. Preprocesamiento de datos:
  - Se realizó una exploración y limpieza inicial exhaustiva del conjunto de datos.
  - Se imputaron los datos con interpolación lineal
  - Se arreglaron los valores faltantes en las variables categóricas creando una nueva categoría ‘No indica’
  - Se escalizaron los datos según su comportamiento estadístico
  - Se realizó PCA
  - Se seleccionaron las variables de interés, incluyendo las características del suelo y la radiación en diferentes longitudes de onda.
2. Aprendizaje no supervisado con K-means:
  - Se aplicó el algoritmo de K-means para identificar clústeres en función de las características seleccionadas.
  - Se realizó una exploración de los clústeres encontrados y se analizaron las relaciones intrínsecas entre las variables.
  - Se utilizó el método del codo y el coeficiente de silueta para determinar el número óptimo de clústeres
3. Clasificación binaria:
  - Se realizó un clasificador binario utilizando el algoritmo de regresión logística, SVM, y DecisionTrees.

- El objetivo fue determinar a qué clúster pertenecía un tipo de cultivo específico (en este caso, el **CACAO**).
4. Evaluación del modelo:
- Se creo un set de entrenamiento, uno de validación y uno de pruebas.
  - Se realizo validación cruzada para entrenar los modelos
  - No se requirió REGULARIZACIÓN para los propositos.
  - Se utilizaron diversas métricas de evaluación, como precisión, coeficiente de correlación de Matthews (MCC) y a la Matriz de Confusión para evaluar el desempeño del modelo de clasificación.

## Resultados

El análisis realizado permitió obtener los siguientes resultados:

- Se identificaron clústeres basados en las características del suelo y la radiación en diferentes longitudes de onda.
- Se exploraron las relaciones intrínsecas entre las variables dentro de los clústeres.
- Se determinó a qué clúster pertenecía el tipo de cultivo específico del **CACAO** utilizando un clasificador binario.
- Se evaluó el desempeño del modelo de clasificación utilizando métricas como precisión, área bajo la curva ROC y coeficiente de correlación de Matthews (MCC).

## Conclusiones

Finalmente, hemos concluido con el desarrollo del proyecto final de IA despues de analizar un dataset real sobre las propiedades del suelo y la absorbanca en diferentes longitudes de onda.

Mi idea inicial era entender los Fertilizantes que se aplican a los suelos, es decir yo quería analizar como las características de pH, Al, Fe, Mg, Mn, Cu, B, y demas minerales influyen en la caracterización del suelo guiada por el tipo de cultivo y por la ubicación geográfica (Departamento). Es por esto que el análisis inicial realizado con aprendizaje No Supervisado con el algoritmo K-means utilizo las características de interes antes mencionadas junto con la radiación de los diferentes longitudes de onda. El proyecto concluyo estudiando el comportamiento de los 2 clústers que se generaron tras la busqueda. Se intentó explorar un poco los datos y ver relaciones intrínsecas de los clusteres. Finalmente se decidió realizar un clasificado binario con el fin de determinar para un tipo de cultivo específico que fue ‘CACAO’ a que cluster de los hallados previamente pertenecian.

Esto puede servir para comprender las características del suelo que se relacionen con un cluster en particular y por eso identificar a que tipo de cluster pertenecen los cultivos es de gran utilidad para indagar y conocer más sobre las propiedades adecuadas de los suelos.