

# ANÁLISIS DE CARACTERÍSTICAS DEL SUELO PARA LA CLASIFICACIÓN DE CULTIVOS AGRÍCOLAS

William Andrés Gómez Roa  
Pontificia Universidad Javeriana, Bogotá, Colombia  
Inteligencia Artificial - Proyecto Final

In this project, we conducted a comprehensive analysis of soil properties and absorbance at different wavelengths to classify agricultural crops. Unsupervised learning techniques were employed, and a binary classifier was applied to determine the cluster to which the cocoa crop belongs.

**Index Terms**—SoilProperties, UnsupervisedLearning, ClusterAnalysis, MachineLearning, AgriculturalCrops, Absorbance, Fertilizer.

## I. INTRODUCCIÓN

En este proyecto, se realizó un análisis exhaustivo de los datos de propiedades del suelo y se aplicaron técnicas de aprendizaje automático. Se exploraron diferentes algoritmos supervisados y no supervisado, como el algoritmo K-means, para identificar patrones y agrupar los datos en diferentes categorías. Además, se implementaron clasificadores binarios utilizando para determinar a que clúster pertenecen los cultivos de cacao. Se evaluaron diversas métricas de desempeño para medir la calidad de los modelos y se realizaron análisis estadísticos para examinar las relaciones entre las variables del suelo y los cultivos.

## II. DESARROLLO

Durante el desarrollo del proyecto, se realizó una exploración y limpieza exhaustiva de los datos, incluyendo la imputación de valores faltantes y la escalización de los datos. A continuación, se aplicó el análisis de componentes principales (PCA) para reducir la dimensionalidad y seleccionar las variables de interés relacionadas con las propiedades del suelo y la radiación. Posteriormente, se implementó el algoritmo de K-means para identificar clústeres basados en las características seleccionadas. Utilizando el método del codo y el coeficiente de silueta, para determinar el número óptimo de clústeres. Luego, se realizó una clasificación binaria utilizando algoritmos como regresión logística, SVM y árboles de decisión para asignar los cultivos agrícolas a los clústeres identificados, centrándonos específicamente en el cultivo de cacao. Por último, se evaluó el modelo utilizando conjuntos de entrenamiento, validación y pruebas, validación cruzada y se utilizaron métricas como precisión, coeficiente de correlación de Matthews (MCC) y la matriz de confusión para evaluar el rendimiento del modelo de clasificación. No fue necesario aplicar regularización. El proyecto se realizó en un 'jupyter notebook' y se subió a Github junto con el dataset, los objetivos iniciales del proyecto y un video explicativo del mismo.

## III. RESULTADOS

### A. Agrupamiento por Clusters utilizando k-means

Se utilizaron los datos de pH, Al, Mg, Mn y los demás minerales del suelo junto con la Absorbancia para dividir el dataset en 2 clusters. Luego se observó la distribución de los cultivos de 'CACAO' en dichos clusters como se observa a continuación.

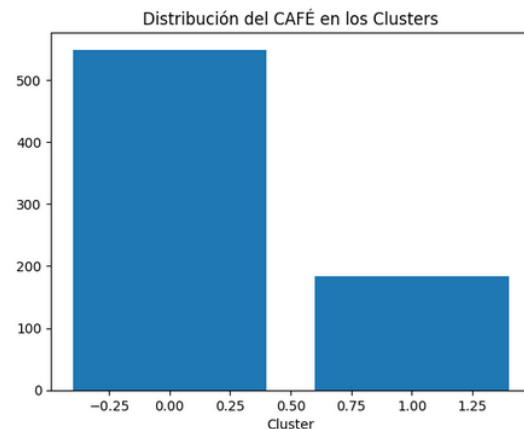


Fig. 1. Distribución de los cultivos de Cacao en los 2 clusters encontrados

Así mismo se realizó la clasificación supervisada de los cultivos de cacao en los 2 clusters previamente mencionados y se evaluaron las métricas que se observana continuación.

```
Logistic Regression:
Validation Accuracy: 0.9983818770226537
Validation MCC: 0.9966244076292227

Decision Trees:
Validation Accuracy: 0.9700647249190939
Validation MCC: 0.9374130657735569

Support Vector Machines (SVM):
Validation Accuracy: 0.9943365695792881
Validation MCC: 0.988179217693462

Average Cross-Validation Accuracy (Logistic Regression): 0.9975715777106045
Average Cross-Validation Accuracy (Decision Trees): 0.9819228768583287
Average Cross-Validation Accuracy (Support Vector Machines): 0.991095299933433
```

Fig. 2. Métricas de los clasificadores binarios entrenados

Todo se realizó en Jupyter-Notebook

#### IV. CONCLUSION

Finalmente, hemos concluido con el desarrollo del proyecto final de IA despues de analizar un dataset real sobre las propiedades del suelo y la absorbancia 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 clusters que se generaron tras la busqueda. Se intentó explorar un poco los datos y ver relaciones intrínsecas de los clusters. 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.