

IDENTIFICACIÓN DE LOS PARÁMETROS ÓPTIMOS DE FERTILIZACIÓN DEL SUELO MEDIANTE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO PARA LA PRODUCCIÓN AGRÍCOLA EN COLOMBIA

El conjunto de datos con el que se trabajará en este proyecto final es una base de datos espectrales y las características del suelo para la producción agrícola en Colombia. Los datos incluyen información sobre las propiedades físicas y químicas del suelo, así como datos espectrales, geográficos y descriptivos del cultivo como su nivel de productividad, además contiene la recomendación de algún experto de si se requiere utilizar fertilizante en dicha muestra de suelo según su cultivo.

El análisis de los datos espectrales permite obtener información sobre la composición química del suelo, mientras que la información descriptiva del cultivo y la recomendación sobre la necesidad de fertilizantes son importantes factores a tener en cuenta en la producción agrícola.

La identificación de los valores de nutrientes óptimos del suelo que permitan que la productividad de un cultivo sea la máxima son de gran interés para la preparación y aplicación correcta de los fertilizantes en dichos terrenos.

OBJETIVOS:

- Realizar un análisis exploratorio de los datos para familiarizarse con el *dataset* e identificar problemas de des balance, así como problemas generales con los datos, además se realizará una limpieza y preprocesamiento de los datos para que estén en un formato numérico adecuado, además llevar a cabo normalización, reducción de dimensionalidad (PCA), y la eliminación de valores atípicos, con el fin de obtener una base de datos de alta calidad y lista para aplicar a técnicas de aprendizaje automático.
- Se utilizará la técnica de aprendizaje no supervisado, *clustering*, para agrupar los suelos en diferentes categorías basadas en su espectro y los niveles de nutrientes. Luego se analizarán las características de los *clusters* resultantes para determinar los patrones en la distribución geográfica y el tipo de cultivo e identificar así los *clusters* que tienen una producción buena y están bien fertilizados. Por último, se quiere hacer una regresión para determinar cuáles son los mejores parámetros nutricionales del suelo según el tipo de cultivo y la ubicación geográfica (Departamento).