Especialización en analítica de datos Est. Pedro Bryan Congo Santander Año 2.023

Visualizadores y Pipelines de Procesamiento de datos ACA3



14789 **14720**

 \blacksquare



PROBLEMA A RESOLVER

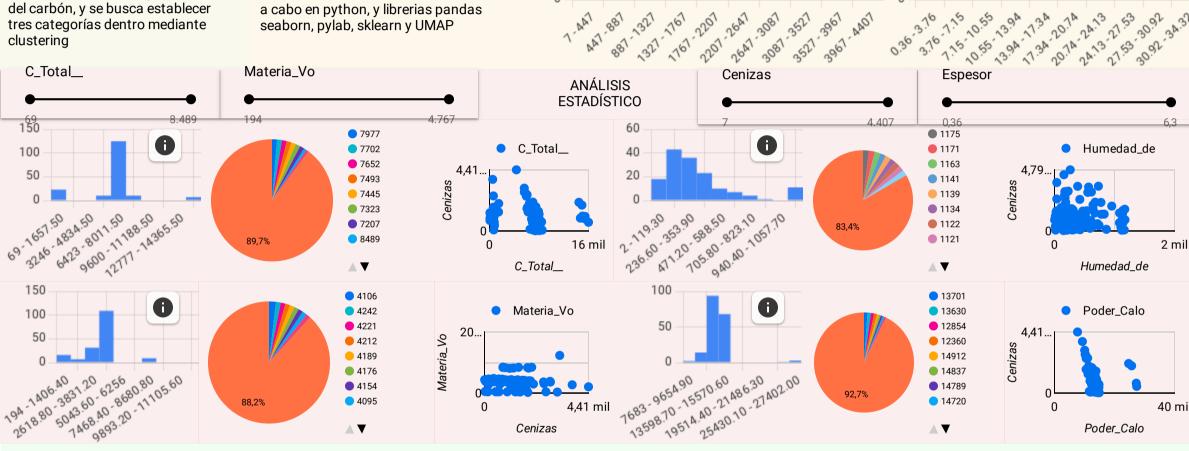
Se trabajo con una base de datos donde se encuentran los resultados y mediciones de calidad y de forma del carbón, y se busca establecer tres categorías dentro mediante clustering

ENTRENAMIENTO DEL MODELO

Se tomaron las variables de cenizas y espesor para llevar a cabo el entrenamiento del modelo. esto se llevo a cabo en python, y librerias pandas seaborn, pylab, sklearn v UMAP





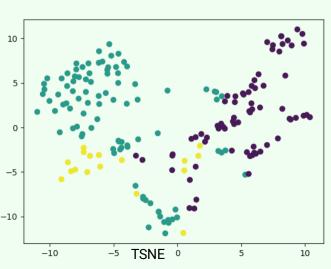


4,41 mil

Cenizas

HIPÓTESIS

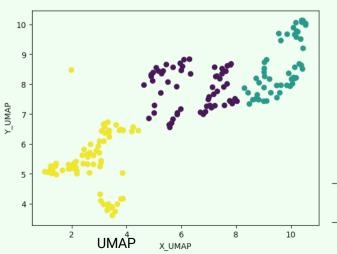
Se diseño la hipótesis de que se podía categorizar en tres categorías elementos, se trabajo con el modelo PCA, el modelo TSNE y el modelo UMAP. este ultimo permitió generar una mejor categorización

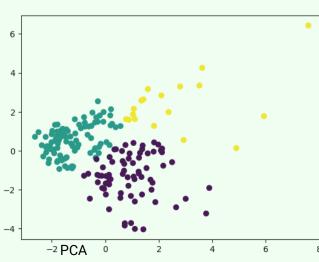


4154

4095

 \blacksquare





0

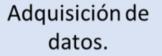
Poder_Calo

40 mil

Visualizadores y Pipelines de Procesamiento de datos ACA 3







Leer los datos de las muestras de carbón desde un archivo CSV.

Almacenamiento de Datos.

Crear una conexión a la base de datos SQLite para almacenar los datos procesados en una tabla.

Procesamiento de Datos.

Calcular el promedio de calidad para cada muestra.

Análisis de Datos.

Realizar una consulta SQL para recuperar muestras de alta calidad según un umbral específico.

Almacenamiento calidad 1.

Se almacena la muestra con calidad tipo 1

Almacenamiento calidad 2.

Se almacena la muestra con calidad tipo 2

Almacenamiento calidad 3.

Se almacena la muestra con calidad tipo 3