

Trabajo Practico N° 1

Minería de datos

Objetivo

El objetivo de este trabajo practico es integrar los conocimientos adquiridos en las unidades 2 y 3 en un problema real asociado a los cultivos.

Actividades

1. Descargar un conjunto de datos, Crop_recommendation.csv¹, para realizar el trabajo práctico.
2. Analizar los atributos del conjunto de datos (distribuciones, valores, outliers, tipos de datos, etc.) y elegir un método de estandarización.
3. Realizar PCA y determinar el número de componentes principales considerando alguno de los 3 criterios datos en la práctica. Graficar la varianza acumulada y las componentes de PCA en un grafico 2 o 3D con sus respectivas clases.
4. Aplicar Isomap y analizar los resultados obtenidos variando el numero de vecinos y componentes. Realizar un grafico en 2D de utilizando dos componentes.
5. Aplicar t-SNE y analizar los resultados obtenidos variando el número de iteraciones, componentes y perplejidad. Realizar un gráfico en 2D de utilizando dos componentes.
6. Aplicar K-means y analizar los resultados obtenidos variando el número de clusters y obtener el número óptimo de clusters mediante GAP. Realizar un gráfico en 3D de utilizando tres atributos de los datos y donde los colores estén asociados a los clusters.
7. Aplicar clustering jerárquico y determinar cuál número sería el que mejor represente los datos. Utilizar el score de Silhouette y calcular el número óptimo de cluster por medio de GAP.

Presentación

La entrega es por grupos de dos estudiantes y se entregan dos archivos por grupo. Cualquier integrante del grupo puede hacer la entrega mediante el campus de la materia.

El formato del informe deberá ser pdf; mientras que el código deberá ser py.

El informe deberá tener una carátula en la que se indique: año, materia, integrantes. Además, deberá contar con una sección de conclusiones al final del mismo.

¹ <https://www.kaggle.com/datasets/chitrakumari25/smart-agricultural-production-optimizing-engine>

Trabajo Practico N° 1

Minería de datos

Las entregas fuera del plazo establecido no serán consideradas salvo excepciones previamente justificadas por el grupo.