

Instrucciones

- Puede utilizar las herramientas disponibles para resolver los problemas identificados en cada uno de los puntos del examen, con excepción de comunicación con entes externos y el uso de inteligencia artificial.
- Utilice Python/Julia para resolver los problemas definidos.
- El examen tiene una duración de dos horas.

1. Clustering/Agrupamiento

Usando los datos de 'cars.csv' realice lo siguiente:

- Análisis univariante
- Análisis bivariante
- Normalizar los datos (MinMaxScaler)
- PCA en R^2
- Definir la cantidad ideal de clusters usando KMeans. Para esto se debe mostrar visualizaciones de la inercia y el coeficiente de silueta. No es necesario simular con diferentes métricas,
- Usando la misma cantidad de clusters definida en el numeral anterior, genere un modelo usando Hierarchical Clustering de tipo aglomerativo, recuerde realizar el dendograma. Luego muestre los resultados de ambos modelos usando PCA en R^2 . ¿Los resultados son similares?

2. Regresión Lineal Simple

Dataset: El dataset contiene información sobre precios de vivienda con las siguientes variables:

- price: precio de la vivienda (target)

- area: área de la vivienda en pies cuadrados
- bedrooms: número de habitaciones
- bathrooms: número de baños
- stories: número de pisos
- parking: número de plazas de parking

Tareas a realizar

1. Análisis exploratorio de datos

- Realice un análisis estadístico descriptivo de todas las variables.
- Genere gráficos de dispersión entre la variable objetivo (price) y cada variable independiente.
- Calcule la matriz de correlación e identifique las variables más correlacionadas con el precio.
- Detecte y maneje valores atípicos si es necesario.

2. Regresión lineal simple

- Seleccione la variable más correlacionada con el precio para un modelo de regresión simple.
- Implemente manualmente el algoritmo de regresión lineal simple de scikit-learn.
- Calcule los parámetros del modelo (pendiente e intercepto).
- Evalúe el modelo usando:
 - Error cuadrático medio (MSE).
 - Coeficiente de determinación (R^2)
- Realice predicciones sobre el conjunto de prueba.
- Evalúe el modelo usando:
 - Error cuadrático medio (MSE).
 - Coeficiente de determinación (R^2) .
- Genere un gráfico comparando las predicciones con los valores reales.