

## Ejercicio 01: Predicción de Presión Arterial Sistólica usando Regresión Lineal

**Objetivo:** Desarrollar un modelo de regresión lineal simple y múltiple para predecir la presión arterial sistólica (PA) de pacientes, usando variables como la edad, el índice de masa corporal (IMC) y el nivel de colesterol.

### Instrucciones:

#### 1. Generación de Datos:

- Simula un dataset de 300 pacientes con las siguientes variables:
  - **edad**: entre 18 y 80 años.
  - **imc**: valores entre 18 y 38, siguiendo una distribución normal.
  - **colesterol**: valores entre 130 y 300 mg/dL.
  - **pa\_sistolica**: variable dependiente, relacionada positivamente con edad, IMC y colesterol, más un componente aleatorio.

#### 2. Exploración de Datos:

- Visualiza una muestra de los datos y obtén estadísticas descriptivas.
- Genera histogramas para analizar la distribución de cada variable.
- Crea una matriz de correlación para observar las relaciones entre variables.
- Realiza gráficos de dispersión entre la PA sistólica y cada variable independiente.

#### 3. Regresión Lineal Simple:

- Desarrolla un modelo de regresión lineal simple para cada predictor (**edad**, **imc**, **colesterol**) por separado.
- Divide los datos en conjunto de entrenamiento (70%) y prueba (30%).
- Entrena el modelo y realiza predicciones sobre el conjunto de prueba.
- Evalúa el modelo usando el Error Cuadrático Medio (MSE) y el coeficiente de determinación ( $R^2$ ).
- Visualiza la relación entre los valores reales y predichos para cada modelo.
- Usa una librería estadística para obtener el resumen del modelo y analizar significancia de coeficientes.

#### 4. Regresión Lineal Múltiple:

- Construye un modelo de regresión múltiple usando **edad**, **imc** y **colesterol** como predictores.
- Realiza el mismo procedimiento de división de datos, entrenamiento, predicción y evaluación.
- Obtén y analiza los coeficientes de cada variable.

#### 5. Evaluación de Modelos:

- Compara el desempeño de los modelos simples con el modelo múltiple.
- Determina qué variable tiene mayor impacto en la predicción de la PA sistólica.

#### 6. Visualización de Resultados:

- Grafica los residuos del modelo múltiple para verificar la distribución y homogeneidad.
- Compara gráficamente los valores predichos y reales.

#### 7. Conclusiones:

- Interpreta los coeficientes y la importancia de cada variable.
- Discute las implicaciones clínicas de los resultados obtenidos.

#### Pistas:

- Usa librerías como **pandas**, **numpy**, **matplotlib**, **seaborn**, **scikit-learn** y **statsmodels**.
- Asegúrate de aplicar buenas prácticas de preprocesamiento y visualización.
- Presta atención a la significancia estadística de los coeficientes y la interpretación de los valores  $R^2$  y MSE.
- En la regresión múltiple, observa posibles colinealidades entre variables.

#### Producto esperado: Un informe que incluya:

- Visualizaciones descriptivas.
- Resultados de los modelos (coeficientes, MSE,  $R^2$ ).
- Gráficos de predicciones vs valores reales.
- Análisis de residuos.
- Conclusiones bien fundamentadas.