



# Tecnológico de Monterrey

**Retro 3**

Rafael Mercado Garduño

A01746991

ITESM CEM

Grupo: 101

Instituto Tecnológico y de Estudios Superiores de Monterrey

### **Justificación del dataset:**

He seleccionado el dataset "Calorias FastF.csv" porque contiene variables numéricas directamente relacionadas con el objetivo de predecir si un alimento tiene más de 219 calorías. Las características seleccionadas, como las calorías provenientes de grasas y las grasas saturadas, son adecuadas para el algoritmo `MLPClassifier`, ya que maneja bien relaciones no lineales entre variables y puede aprovechar la naturaleza cuantitativa de estos datos para realizar predicciones precisas.

### **Generalización del modelo:**

Para asegurar que el modelo generaliza correctamente, dividí el dataset en conjuntos de entrenamiento (80%), validación (20%) y prueba (20%). Los resultados obtenidos fueron 0.74 de precisión en entrenamiento, 0.74 en validación y 0.77 en prueba, lo que demuestra que el modelo tiene una buena capacidad de generalización y que no está sobreajustado a los datos de entrenamiento.

### **Sesgo:**

Para diagnosticar el sesgo (bias) de mi modelo, comparé el rendimiento en los conjuntos de entrenamiento y validación. El modelo alcanzó una precisión de 0.74 en ambos conjuntos, lo que indica que el modelo tiene un sesgo bajo, ya que está aprendiendo bien los patrones del dataset. Si hubiera tenido un sesgo alto, las métricas de rendimiento en ambos conjuntos habrían sido consistentemente bajas, lo que indicaría que el modelo no captura bien las relaciones en los datos.

### **Varianza:**

Al analizar el grado de varianza, el modelo mostró una precisión de 0.74 en validación y 0.77 en prueba, lo que sugiere un grado de varianza bajo. Esta pequeña diferencia en el rendimiento entre los conjuntos de validación y prueba indica que el modelo es robusto y no está sobreajustado a los datos de entrenamiento. Una alta varianza habría producido una mayor diferencia entre los resultados de entrenamiento y prueba, indicando sobreajuste.

### **Nivel de ajuste del modelo:**

El rendimiento del modelo muestra que está fit, ya que las métricas de rendimiento son consistentes entre los conjuntos de entrenamiento, validación y prueba. El modelo

no presenta señales de underfitting ni overfitting, ya que no hay diferencias significativas entre los resultados obtenidos en cada conjunto.

## Mejoras:

Para mejorar el rendimiento del modelo, implementé varias técnicas de ajuste. En primer lugar, apliqué regularización L2 mediante el ajuste del parámetro `alpha` del `MLPClassifier`, lo que ayuda a penalizar los coeficientes grandes, previniendo el sobreajuste y mejorando la capacidad de generalización del modelo. Además, ajusté el tamaño de la capa oculta (`hidden\_layer\_sizes`) y la tasa de aprendizaje (`learning\_rate\_init`). Estas modificaciones permitieron optimizar el rendimiento del modelo.

Tras realizar estos ajustes, los resultados mostraron una mejora significativa en el rendimiento del modelo. En el conjunto de validación, la precisión aumentó a 0.81, el recall a 0.78 y el F1 Score a 0.78. Por otro lado, en el conjunto de prueba, la precisión mejoró a 0.83, el recall a 0.83 y el F1 Score a 0.83. Estos cambios reflejan una mejora notable en el equilibrio entre el sesgo y la varianza, asegurando que el modelo generaliza mejor a datos no vistos.

## Gráficas del Modelo



