

# Laboratorio 5 - Autocodificadores

## Objetivos

- Comprender la arquitectura y funcionamiento de un autocodificador.
- Utilizar un autocodificador para la reducción de dimensionalidad en un conjunto de datos reales.
- Interpretar las salidas del modelo para descubrir percepciones nuevas ("insights") sobre el consumo de comida en diferentes países del Reino Unido.

## Herramientas Requeridas

- Python 3.x
- Bibliotecas: TensorFlow , Pandas, Matplotlib, scikit-learn

## Datos

Se proporciona el archivo UK\_foods.csv, que muestra el consumo promedio de 17 tipos de comida en gramos por persona por semana para cada país en el Reino Unido (UK).

La tabla muestra algunas variaciones interesantes en los diferentes tipos de comidas, pero las diferencias generales, o globales, no son tan obvias.

## Partes del Laboratorio

### Parte 1: Preparación de Datos (15 puntos)

- 1.1 Cargue los datos en un DataFrame de Pandas.
- 1.2 Realice una exploración básica de los datos (estadísticas descriptivas).
- 1.3 Normalice los datos, si es necesario.

### Parte 2: Construcción del Autocodificador (40 puntos)

- 2.1 Dada la pequeña cantidad de datos, utilice todas las observaciones para entrenar el modelo.
- 2.2 Defina la arquitectura del autocodificador. Asegúrese de reducir las dimensiones a 2 en la capa

de “cuello de botella”.

2.3 Compile y entrene el modelo.

2.4 Evalúe el rendimiento del autocodificador utilizando medidas como la pérdida de reconstrucción.

### **Parte 3: Reducción de Dimensionalidad y Visualización (35 puntos)**

3.1 Utilice el autocodificador para reducir las dimensiones del conjunto de datos a 2.

3.2 Visualice las dos dimensiones en un gráfico de dispersión.

3.3 Intente interpretar lo que representan estas dos nuevas dimensiones en relación con el consumo de alimentos.

### **Parte 4: Discusión y percepciones nuevas "Insights" (10 puntos)**

4.1 ¿Qué podemos aprender sobre los patrones de consumo de alimentos en los cuatro países del Reino Unido?

4.2 ¿El autocodificador hizo un buen trabajo al reducir la dimensionalidad? ¿Qué métricas de rendimiento apoyan su afirmación?

## **Entregables**

- Código fuente del proyecto.
- Reporte en formato PDF (puede ser en el mismo Notebook) con todas las secciones descritas, incluyendo gráficas y conclusiones.

## **Rúbrica**

### **Parte 1: Preparación de Datos (15 puntos)**

- Carga de datos: 5 puntos
- Exploración básica: 5 puntos
- Normalización de datos: 5 puntos

### **Parte 2: Construcción del Autocodificador (40 puntos)**

- Utilización de todas las observaciones para el entrenamiento: 10 puntos
- Definición de la arquitectura: 10 puntos

- Compilación y entrenamiento: 10 puntos
- Evaluación: 10 puntos

### **Parte 3: Reducción de Dimensionalidad y Visualización (35 puntos)**

- Reducción de dimensionalidad: 15 puntos
- Visualización en gráfico de dispersión: 10 puntos
- Interpretación de dimensiones: 10 puntos

### **Parte 4: Discusión y percepciones nuevas "Insights" (10 puntos)**

- Interpretación de patrones de consumo: 5 puntos
- Evaluación de rendimiento del modelo: 5 puntos

**Total: 100 puntos**