



# Aprendizaje profundo

## Taller EDA – Fase 1

Profesor:

Ing. Julio Omar Palacio Niño, M.Sc.

[palacio\\_julio@javeriana.edu.co](mailto:palacio_julio@javeriana.edu.co)

**Temática:** Proyecto Final Deep Learning

El proyecto final de la asignatura es la oportunidad para explorar un problema de clasificación de Deep Learning en un conjunto de datos del mundo real.

Este taller (Fase 1) constituye la primera entrega oficial del proyecto y se centra en establecer las bases del trabajo. El objetivo es asegurar una correcta definición del problema, un análisis exploratorio de datos (EDA) exhaustivo y un planteamiento metodológico claro antes de proceder con la implementación.

## 1) Definición del Problema y Objetivos

1. **Selección del Dataset:** Seleccionar uno de los datasets propuestos para el proyecto

Red Neuronal Convolutacional:

<https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>

Red Neuronal Recurrente:

<https://www.kaggle.com/datasets/chizhikchi/andalusian-hotels-reviews-unbalanced>

## 2) Análisis del Dataset y Dimensionalidad

2. **Análisis de Dimensionalidad:**

- ⊙ Descripción del tamaño del dataset (número de muestras, número de características).
- ⊙ Para imágenes (CNN): Resolución, número de canales, variabilidad.
- ⊙ Para texto (RNN): Longitud de las reseñas (media, máx, mín), tamaño del vocabulario.
- ⊙

#### **Análisis de la Variable Objetivo (Clases):**

- Identificación clara de la variable a predecir.
- Análisis de balance de clases (distribución gráfica).

#### **Análisis de Características (Features):**

- Descripción de las características (tipo de dato, significado).
- Análisis estadístico (media, mediana, desviación) para datos numéricos.
- Análisis de frecuencia para datos categóricos.
- Visualizaciones relevantes (histogramas, diagramas de dispersión, nubes de palabras, etc.).

#### **Plan de Preprocesamiento:**

- Describir los pasos que se proponen para limpiar y preparar los datos antes de la etapa de modelado.

### **3) Análisis del Dataset y Dimensionalidad**

En esta sección se deben *proponer* (diseñar teóricamente) 2 modelos solicitados en el proyecto final , como parte de la **Metodología**.

**Métricas de Evaluación:** Definir las métricas que se usarán para reportar el rendimiento de los modelos (ej. Accuracy, Precision, Recall, F1-Score, Matriz de Confusión).

### **4) Entregables del Taller**

Para la fecha estipulada, se deben entregar los siguientes dos archivos (desarrollados en Python, Jupyter o Colab ):

- **Reporte Inicial (PDF):** Un documento que contenga el desarrollo completo de las Partes A, B y C de este taller.
- **Código Fuente (Notebook):** Un *notebook* (Jupyter o Colab) que contenga todo el código Python utilizado para el Análisis Exploratorio de Datos (Parte B). El código debe estar limpio, comentado y ser reproducible.