

## Memoria técnica del proceso de modelado Deep Learning

El objetivo de este trabajo es construir un modelo de Deep Learning capaz de predecir el nivel de engagement o la categoría principal de puntos de interés (POIs), combinando información visual (imágenes) y metadatos contextuales.

### Metodología del modelo

#### Descripción

El dataset está compuesto por puntos de interés (POIs) que incluyen tanto imágenes como metadatos (categoría, localización, visitas, likes, bookmarks, etc). El objetivo es predecir una variable objetivo (engagement), combinando estas fuentes heterogéneas mediante un modelo híbrido.

#### Planteamiento inicial

He comenzado analizando el dataset, limpiando y preprocesando las variables (tratamiento de nulos, codificación de variables categóricas y numéricas, verificación de las rutas de imágenes, etc).

#### Aproximaciones previas

He realizadp diferentes notebooks, cada uno suponiendo una iteración sobre el anterior:

##### - *PracticaDL\_Beatriz\_Velayos-1.ipynb*:

Primer intento de modelo híbrido, combinando una red neuronal simple para metadatos y un extractor de características de imágenes (ResNet18). Se encontraron problemas con la preparación del input, la normalización y la integración de ambas ramas. Los resultados iniciales fueron bajos, con dificultades para el entrenamiento conjunto.

##### - *PracticaDL\_Beatriz\_Velayos-2.ipynb*:

Se mejoró la integración de imágenes y metadatos, ajustando la arquitectura y la función de pérdida. Sin embargo, el modelo sufría de overfitting rápido y las métricas de validación eran inestables. Se detectaron problemas de desbalanceo de clases y dificultad para generalizar.

##### -*PracticaDL\_Beatriz\_Velayos-3.ipynb*:

En esta versión se pulieron detalles del preprocesamiento y de la integración del pipeline. El código se hizo más modular y se refinaron métricas e interpretación de resultados. El rendimiento fue más consistente, pero se observaba que para algunas clases el modelo seguía sin funcionar correctamente.

### Descripción del modelo final

El modelo final, desarrollado en PracticaDL\_Beatriz\_Velayos.ipynb, consiste en un pipeline híbrido: una rama basada en ResNet18 preentrenada para procesar imágenes, y otra rama de fully connected layers para los metadatos. Ambas salidas se concatenan y pasan por una capa final para la predicción. Se emplea validación cruzada, métricas detalladas y análisis de errores. El preprocesamiento se ha optimizado y se han aplicado técnicas de regularización y data augmentation para mejorar la generalización. Aunque el accuracy global no es alto, la solución es coherente y sólida respecto a las limitaciones del dataset.

### Resultados obtenidos

El modelo muestra un rendimiento global bajo (accuracy 0.36, F1 ponderado 0.35). Para recomendar o clasificar POIs por su categoría principal, sería poco útil para las clases minoritarias: estos POIs nunca se reconocen, lo que supone un riesgo de dejar fuera recomendaciones relevantes o estratégicas, especialmente en categorías con pocos ejemplos. Para las clases frecuentes ("Ocio", "Cultura", "Escultura", "Historia"), el modelo tiene cierta capacidad para diferenciarlas, pero el rendimiento sigue siendo modesto (F1 entre 0.24 y 0.47). Por tanto, el modelo puede aportar valor solo como ayuda complementaria, pero está lejos de ser fiable como sistema de clasificación automática.

### Conclusiones

El trabajo ha permitido desarrollar un pipeline funcional de Deep Learning multimodal, capaz de integrar imágenes y metadatos para la clasificación de POIs. A lo largo del proceso se han explorado diferentes enfoques, afrontando problemas típicos como el desbalanceo, la integración de datos heterogéneos y la regularización del modelo. El resultado final, aunque lejos de la perfección, constituye una base sobre la que seguir investigando.

### Posibles mejoras y líneas futuras

- Probar arquitecturas de imagen más avanzadas (ResNet50, EfficientNet) o técnicas de fine-tuning más agresivas.
- Aplicar técnicas de oversampling/undersampling o focal loss para mitigar el desbalanceo de clases.
- Explorar modelos de atención (attention pooling) para fusionar mejor las ramas de imagen y metadatos.
- Investigar la explicación de predicciones (ej. Grad-CAM) y análisis de errores para depurar la clasificación de clases minoritarias.
- Recopilar más datos o aplicar técnicas de generación de datos sintéticos para aumentar la representación de clases poco frecuentes.