

PROYECTO DEEP LEARNING ALEJANDRO RIOS SILVA

Con el proyecto hemos conseguido desarrollar un modelo de Deep Learning híbrido capaz de predecir el nivel de engagement de puntos de interés turísticos, integrando información tanto visualmente como en metadatos. Con esto podemos conseguir optimizar estrategias de contenido y conseguir insights valiosos para la toma de decisiones.

Exploración de Datos

Durante la fase de análisis exploratorio, he localizado lo siguiente:

- **Registros totales:** 1,589.
- **Variables:** 7 características principales.
- **Valores nulos:** No se identificaron datos faltantes.
- La variable tags estaba almacenada como tipo object por lo que fue transformada para su correcto procesamiento en el modelo.

He definido el engagement score cogiendo los likes – dislikes + bookmarks.

Esto fue seleccionado como la mejor representación del compromiso de los usuarios con los puntos de interés.

La distribución del engagement score ha mostrado una alta concentración en valores cercanos a 0, con algunos outliers destacados. Después he dividido el dataset en los siguientes subconjuntos:

- **70 % para entrenamiento**
- **15 % para validación**
- **15 % para prueba**

Arquitectura del Modelo

El modelo híbrido consiste en un **ResNet50 preentrenado** en ImageNet para la extracción de características visuales de las imágenes y una red **fully connected** para procesar los metadatos.

1. **ResNet50:**
 - Extracción de características visuales mediante un modelo preentrenado.
2. **Red fully connected:**
 - Uso de Dropout para prevenir sobreajuste.
 - Representación numérica de los tags utilizando embeddings.
3. **Regularización:**

- Aplicación de técnicas como Dropout, Early Stopping y reducción de la tasa de aprendizaje al detectar estancamiento.
4. **Entrenamiento:**
- Función de pérdida: CrossEntropyLoss.
 - Optimizador: Adam con peso de decaimiento.
 - Número de epochs: 50.
-

Resultados del Entrenamiento

- Reducción rápida de la pérdida en las primeras epochs, con una estabilización posterior.
 - Buen equilibrio entre las pérdidas de entrenamiento y validación.
 - La precisión de validación superó a la de entrenamiento en varias epochs, indicando que el modelo generalizó bien.
-

Evaluación del Modelo

Al analizar la matriz de confusión destaco los siguientes puntos:

- **Falsos negativos:** 15 casos.
- **Precisión general:** 92 %.

Interpretaciones:

- El modelo logra una **precisión alta** en la predicción del nivel de engagement para la mayoría de los puntos de interés turísticos.
 - Se obtuvo un buen equilibrio entre las métricas de precisión y recall.
 - Sin embargo, el desempeño en la clase minoritaria (High Engagement) podría mejorarse.
-

Conclusión

El modelo logra una precisión general alta y es efectivo en la predicción de engagement para la mayoría de los POIs. Conseguimos una precisión global del 92 %, lo cual está bastante bien y hemos conseguido tener un buen equilibrio entre las métricas de precisión y recall. Aunque no he conseguido que el desempeño de la clase minoritaria (High Engagement) sea mejor de lo que ha acabado siendo, en definitiva me parece que ha quedado un buen modelo.