

Desarrollo del Modelo y Análisis de Resultados

1 Objetivo

Entrenar y ajustar los modelos utilizados para la detección de ingredientes y la recomendación de recetas, obteniendo resultados accionables a partir de los datos procesados.

2 Ajuste del Modelo de Detección de Ingredientes

Dado que **Grounding DINO** es un modelo preentrenado, no se realizó un entrenamiento desde cero. En su lugar, se propuso realizar un **fine-tuning** utilizando el dataset **COCO**, con el objetivo de mejorar la detección de ingredientes en imágenes específicas del dominio de recetas. Esto no se llevará a cabo en este proyecto ya que no existen los recursos necesarios.

2.1 Comparación con Otros Modelos

Se realizaron pruebas con otros modelos de visión por computadora para evaluar sus capacidades en la detección de ingredientes:

- **CLIP**: Su enfoque multimodal permite relacionar imágenes y texto de manera efectiva, pero su rendimiento en la detección de ingredientes fue inferior. La principal desventaja de este modelo es la ausencia de encuadres, lo cual hace que comprobar la calidad de la detección sea complicado. Además, CLIP no soporta clasificación multimodal, lo cual es esencial en este proyecto.
- **Owl ViT**: Modelo de visión transformer orientado a la detección de objetos, que mostró cierto potencial, pero con menor precisión en la identificación de ingredientes específicos en comparación con Grounding DINO. En muchas pruebas, Owl ViT no fue capaz de identificar ningún objeto en la imagen, lo que limita su aplicabilidad en este contexto.
- **Grounding DINO**: Se confirmó como la mejor opción debido a su capacidad para detectar ingredientes individuales sin necesidad de etiquetas de entrenamiento específicas.

2.2 Optimización y Ajustes

- Se consideró la posibilidad de ajustar hiperparámetros como el umbral de confianza para la detección de ingredientes.
- Se evaluó el desempeño visualmente mediante ejemplos de predicción, observando que el modelo tiene mejor rendimiento en ingredientes comunes, probablemente debido a su mayor representación en el dataset original.

- Se identificó que la detección de ingredientes menos frecuentes es más errática, lo que justifica la necesidad de un dataset específico para mejorar el modelo en este aspecto.

3 Evaluación de Resultados

Dado que no se realizó un entrenamiento tradicional, no se calcularon métricas como precisión, recall o F1-score. En su lugar, se evaluaron los resultados de forma cualitativa:

3.1 Evaluación de Grounding DINO

- Se observó que el modelo tiene un **buen desempeño** en ingredientes comunes como tomate, manzana o plátano.
- En ingredientes menos representados en datasets generales, la detección es más inconsistente.
- No se obtuvieron métricas cuantificables debido a la falta de etiquetas de referencia adecuadas.

3.2 Evaluación del Sistema de Recomendación

- No existen métricas estandarizadas para evaluar la calidad de una recomendación.
- Se compararon diferentes enfoques para la representación de textos:
 - **CountVectorizer**: Se descartó porque trata todos los términos por igual, sin diferenciar palabras comunes de términos más representativos.
 - **Embeddings**: No se utilizaron debido a la necesidad de un modelo entrenado previamente en un dominio similar, lo cual no siempre es viable para ingredientes específicos. Además, los ingredientes pueden tener lexemas similares sin referirse al mismo elemento, lo cual generaría ruido en el sistema de recomendación.
 - **TF-IDF**: Se seleccionó porque penaliza ingredientes demasiado frecuentes, permitiendo que las recomendaciones sean más personalizadas y útiles para el usuario.
- El sistema de **matching bidireccional** mejoró la precisión en la recomendación de recetas basadas en ingredientes disponibles. Este incluye una forma de priorizar aquellas recetas que mejor se ajuste a la disponibilidad de ingredientes y teniendo en cuenta la limitación en cuanto a cantidad de éstos.

4 Aplicación de los Resultados e Insights de Negocio

El sistema desarrollado tiene aplicaciones tanto a nivel de usuario final como en modelos de negocio B2B.

4.1 Aplicación para Usuario Final (Webapp)

- **Identificación automática de ingredientes** a partir de una imagen tomada por el usuario.
- **Recomendación de recetas** con base en los ingredientes detectados, priorizando aquellas que se pueden hacer con lo que el usuario ya tiene. Se incluyen los detalles de las recetas, como los macronutrientes y raciones.
- **Conexión con supermercados** representada en links directos a la compra de productos, además de mostrar los precios de cada uno de los ingredientes necesarios para cada receta (fuente Mercadona).

A partir de estos resultados, se pueden extraer las siguientes recomendaciones estratégicas, aplicables en diferentes contextos. Por ejemplo, la integración de la detección de ingredientes en una webapp permitiría a los usuarios escanear su despensa y recibir recetas en tiempo real, mejorando su experiencia culinaria. Para el sector B2B, la compartición de datos con supermercados podría facilitar estrategias de marketing personalizadas, mientras que la inclusión de enlaces de compra directa aumentaría la conversión de ventas en línea, aunque esto quedará para futuras mejoras.

5 Conclusiones y Próximos Pasos

Este análisis ha permitido evaluar el desempeño de los modelos empleados y establecer mejoras futuras. Como próximos pasos se sugiere:

- Implementar **fine-tuning** de Grounding DINO en una plataforma con GPUs externas.
- Incorporar etiquetas de referencia para evaluar métricas cuantitativas en detección de ingredientes.
- Explorar nuevos enfoques para evaluar la calidad de las recomendaciones del sistema basado en NLP.