

LITCOIN NPL CHALLENGE

Resultados Finales del LitCoin NPL Challenge



PROBLEMA

El reto consiste en analizar los títulos y resúmenes de artículos biomédicos para identificar entidades como genes, enfermedades o productos químicos, y predecir las relaciones entre esas entidades, como asociaciones, correlaciones o interacciones. Los participantes deben entrenar modelos de procesamiento de lenguaje natural (NLP) para realizar esta tarea de forma automática, utilizando los datos proporcionados en formato CSV para entrenar y evaluar sus modelos.



LOS DATOS

0	id	73910	non-null	int64
1	abstract_id	73910	non-null	object
2	type_x	73910	non-null	object
3	entity_1_id	73910	non-null	object
4	entity_2_id	73910	non-null	object
5	novel	73910	non-null	object
6	abstract	73910	non-null	object
7	title	73910	non-null	object
8	entity_ids_x	73910	non-null	object
9	entity_1_mention	73910	non-null	object
10	type_y	73910	non-null	object
11	entity_ids_y	73910	non-null	object
12	entity_2_mention	73910	non-null	object
13	entity_2_type	73910	non-null	object



EL MODELO UTILIZADO

Transformers

Por qué **TRANSFORMERS**



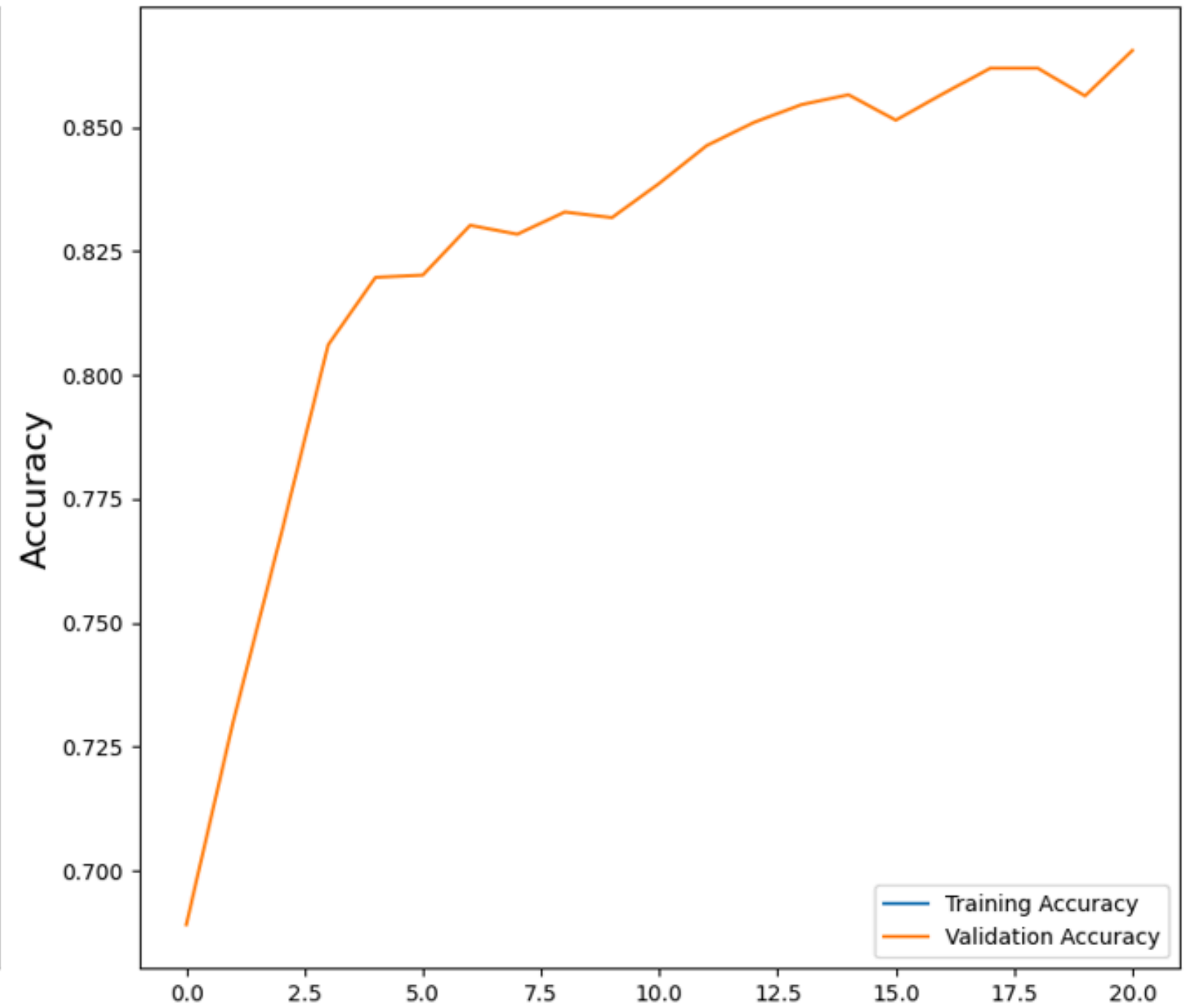
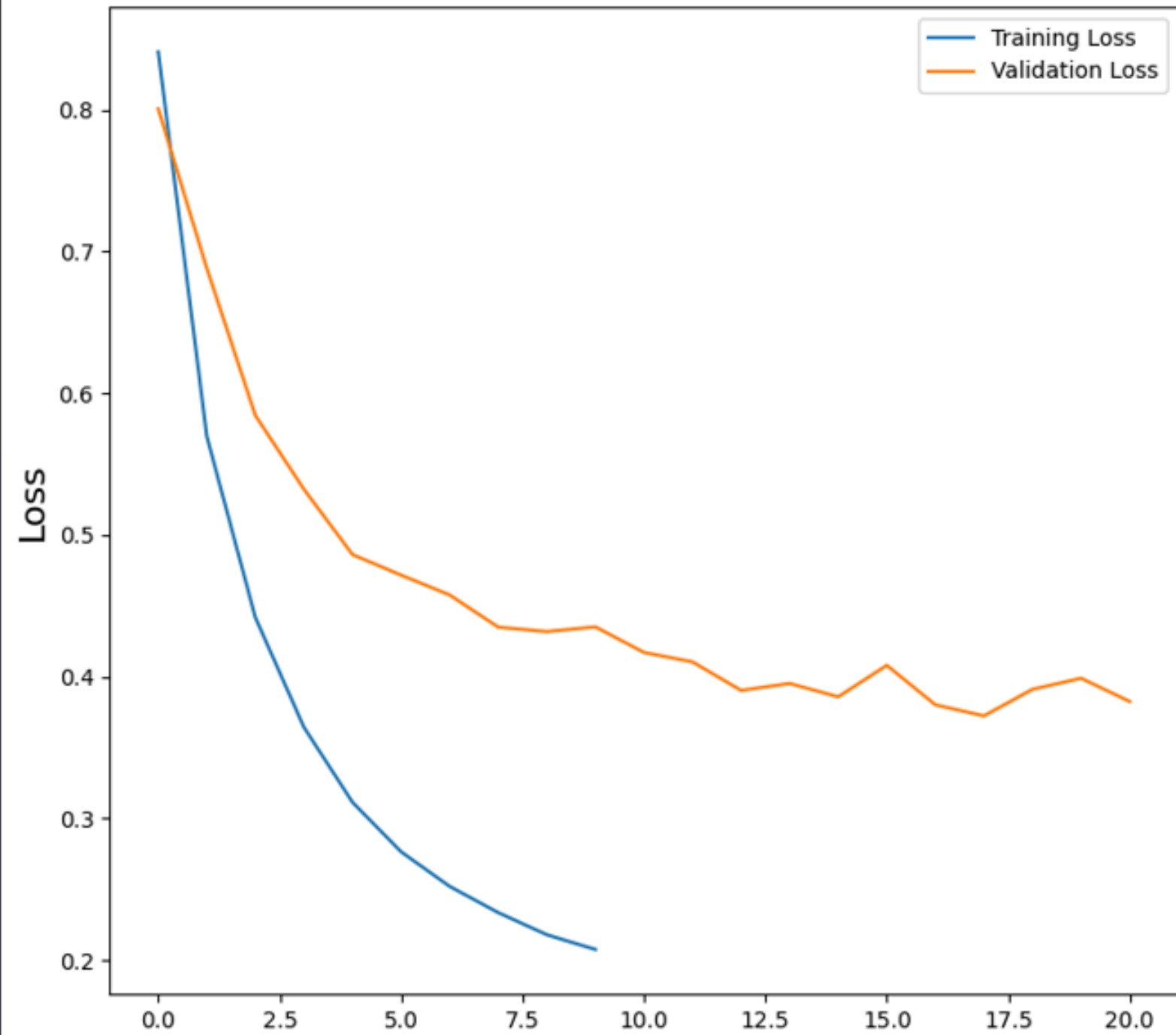
**Optimizado para datos
secuenciales**



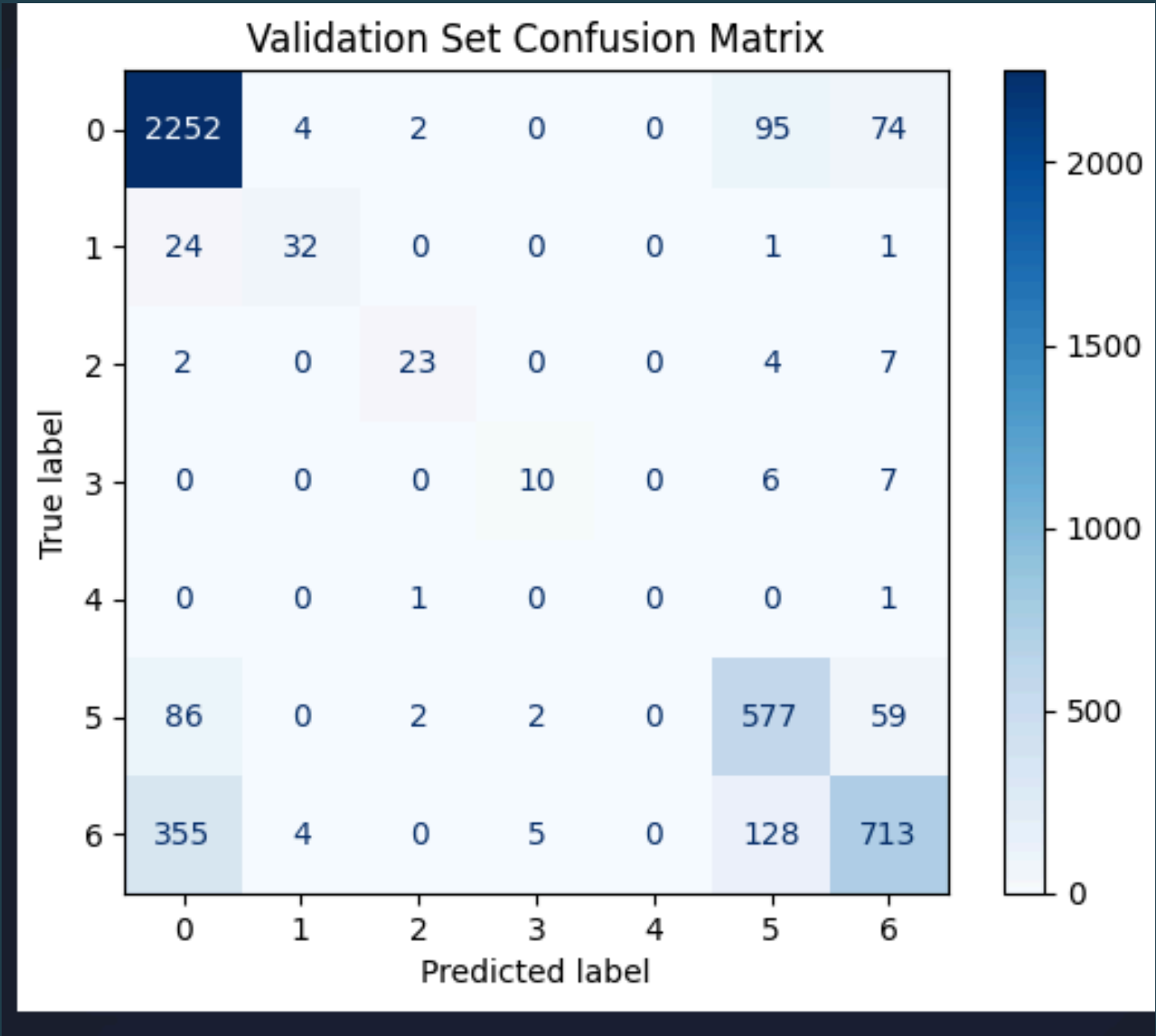
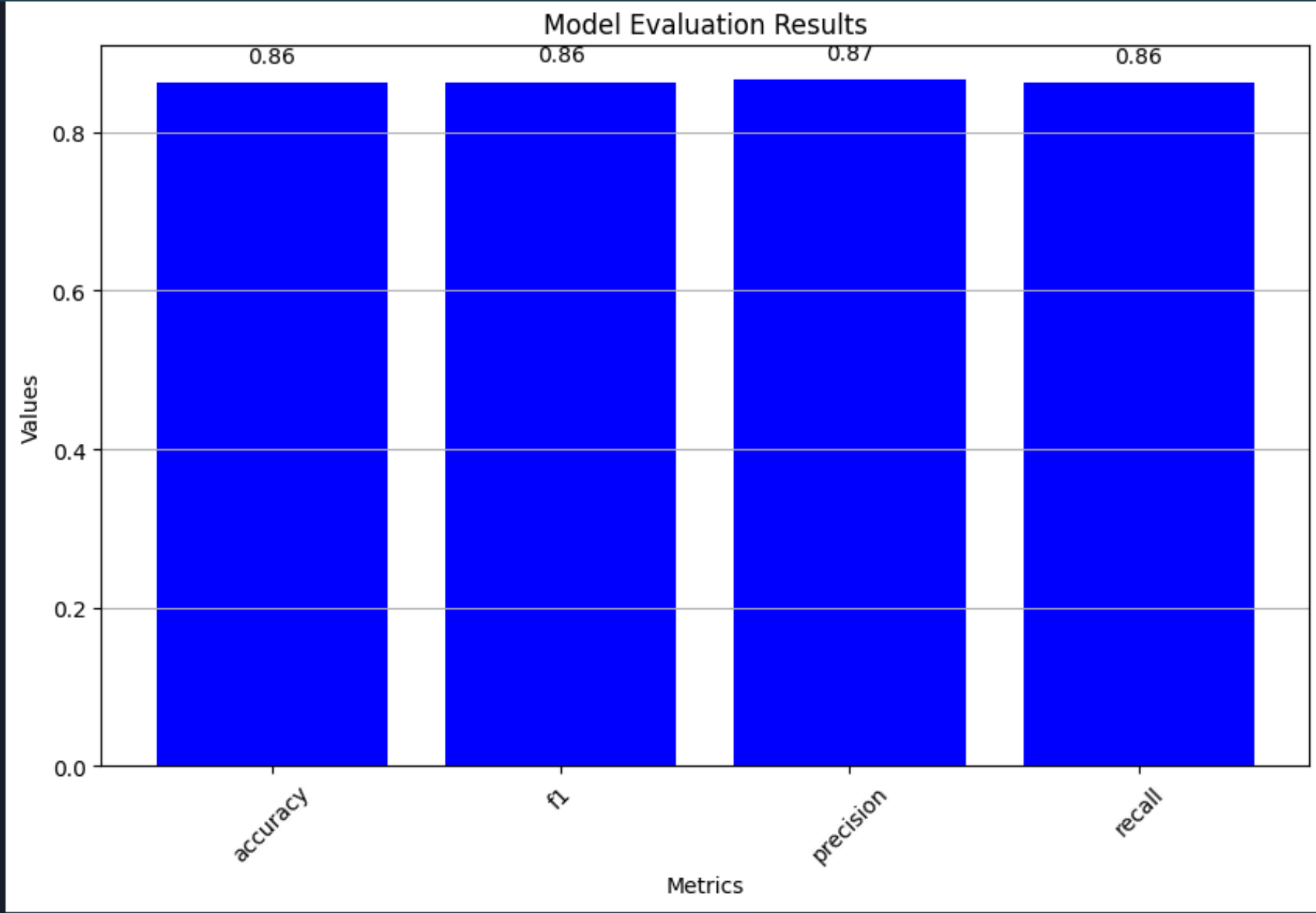
**Mecanismos de atención para
enfocar en las partes más
relevantes**



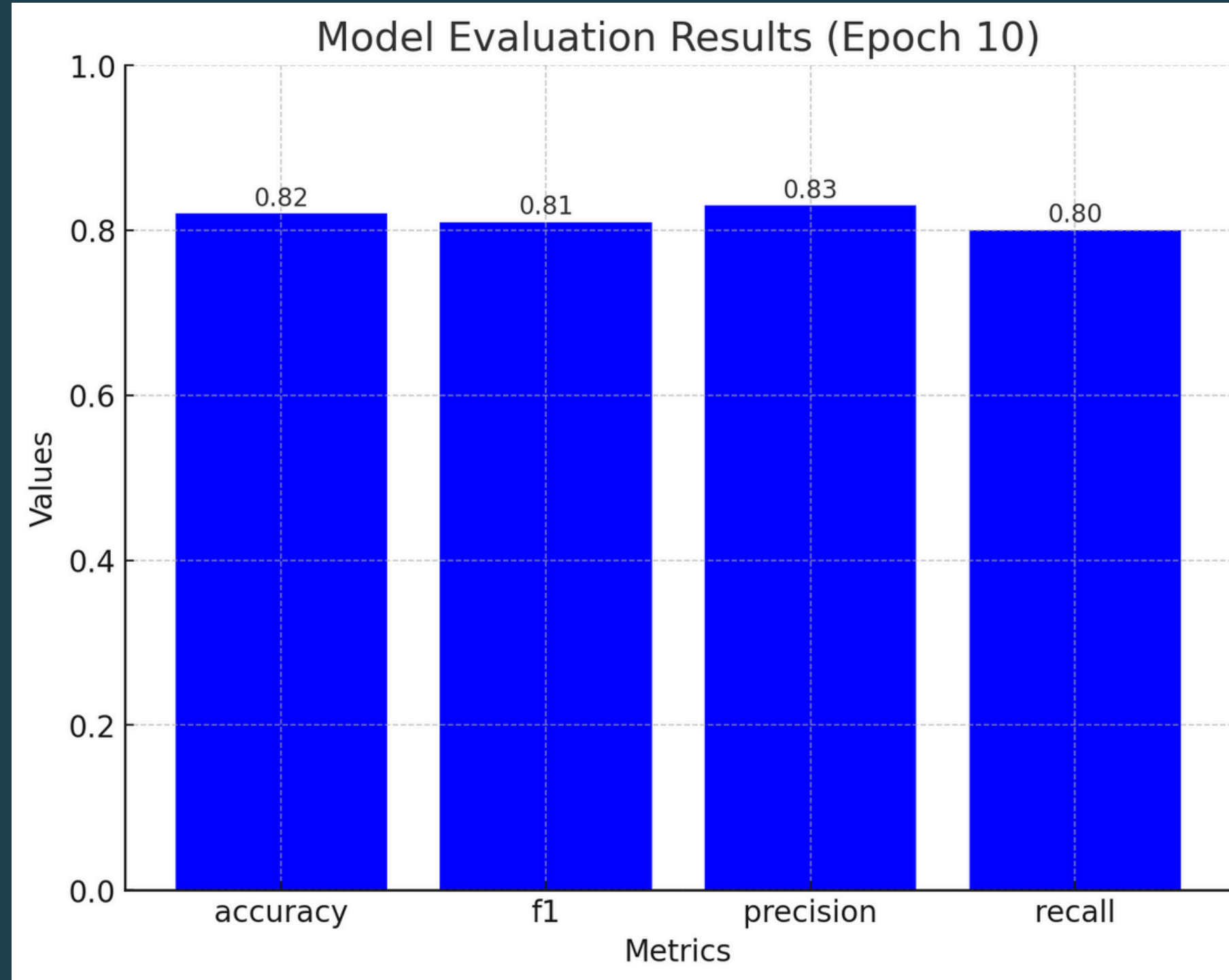
Optimizer : Adam



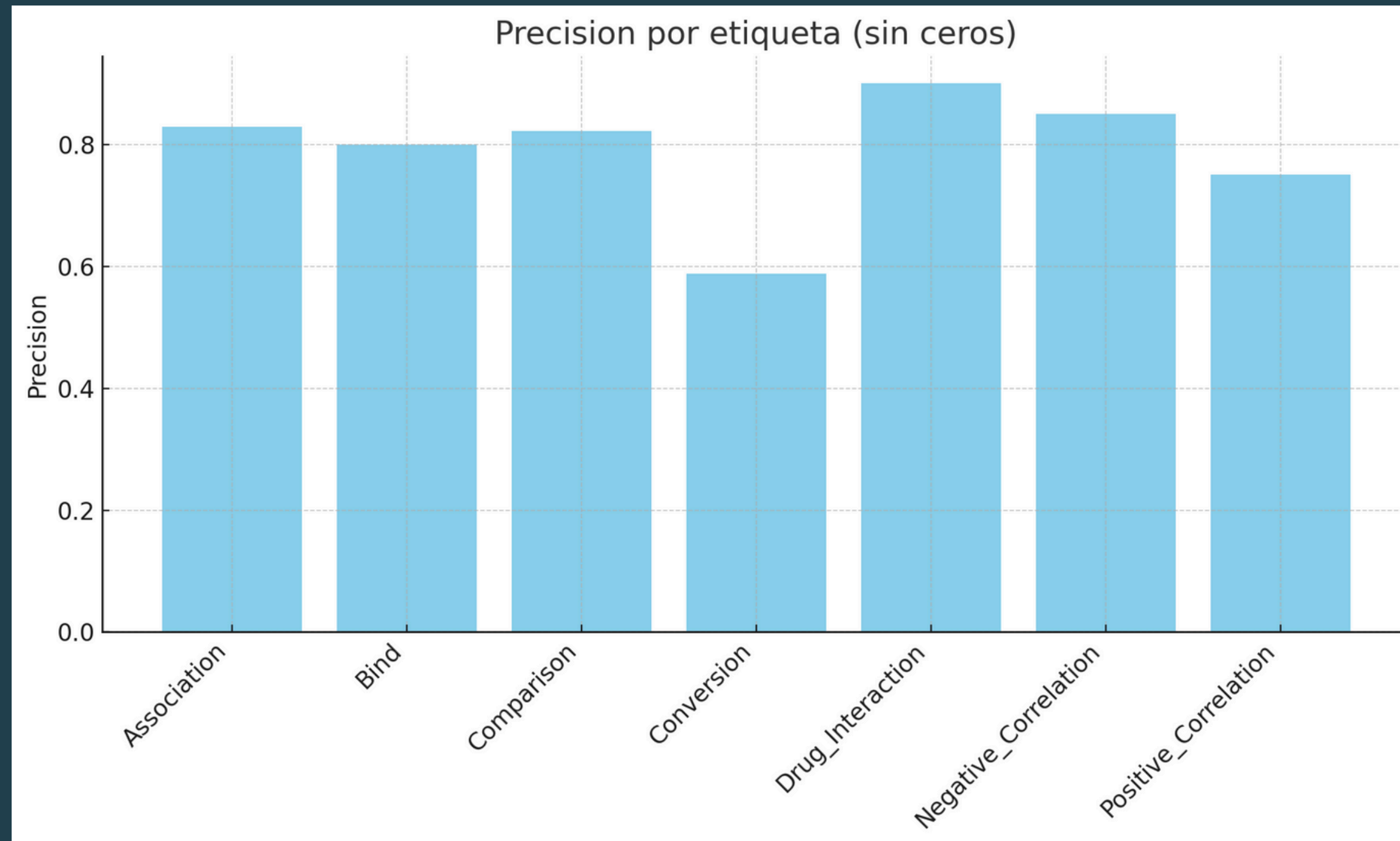
RESULTADOS GLOBALES



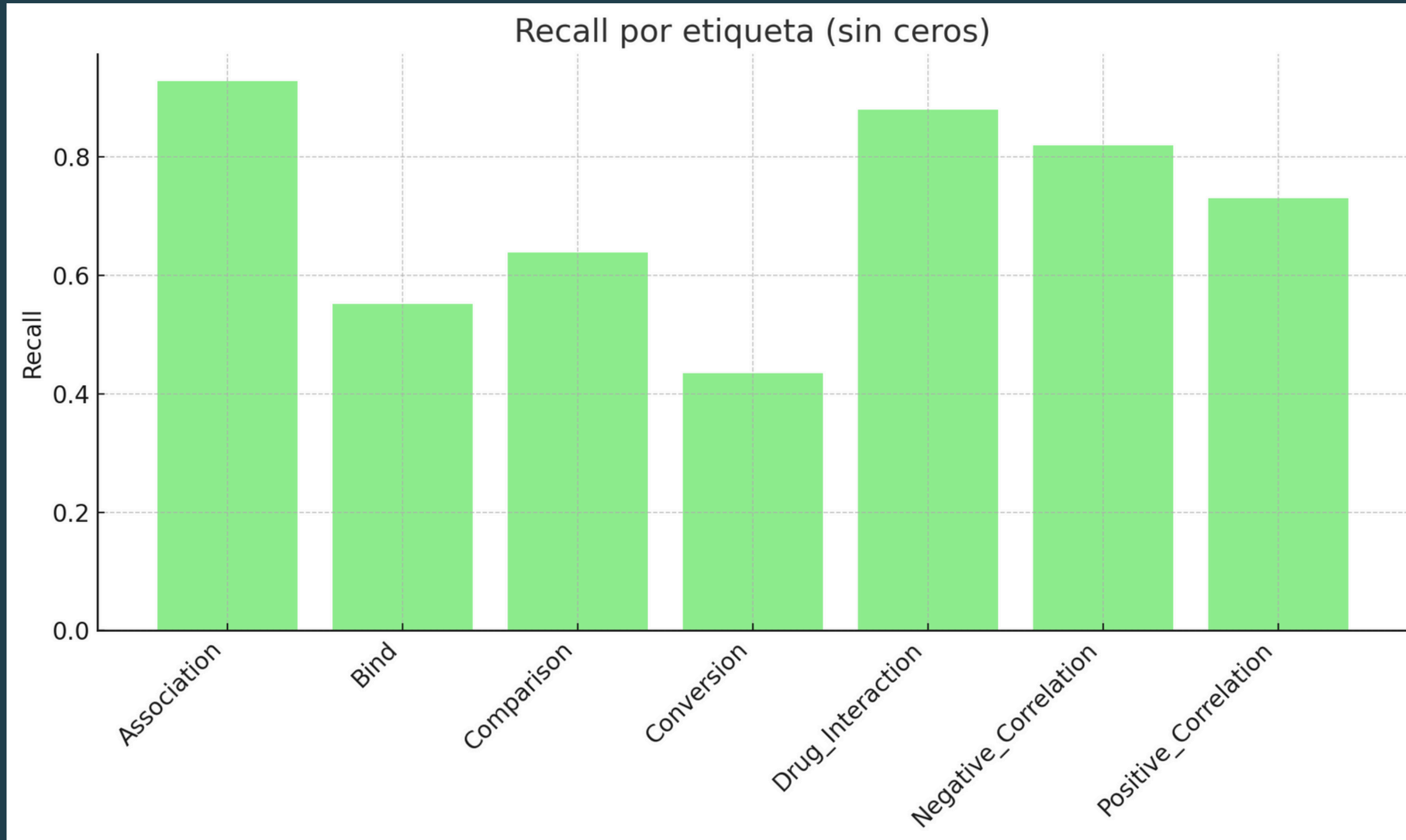
RESULTADOS GLOBALES 10 EPOCAS



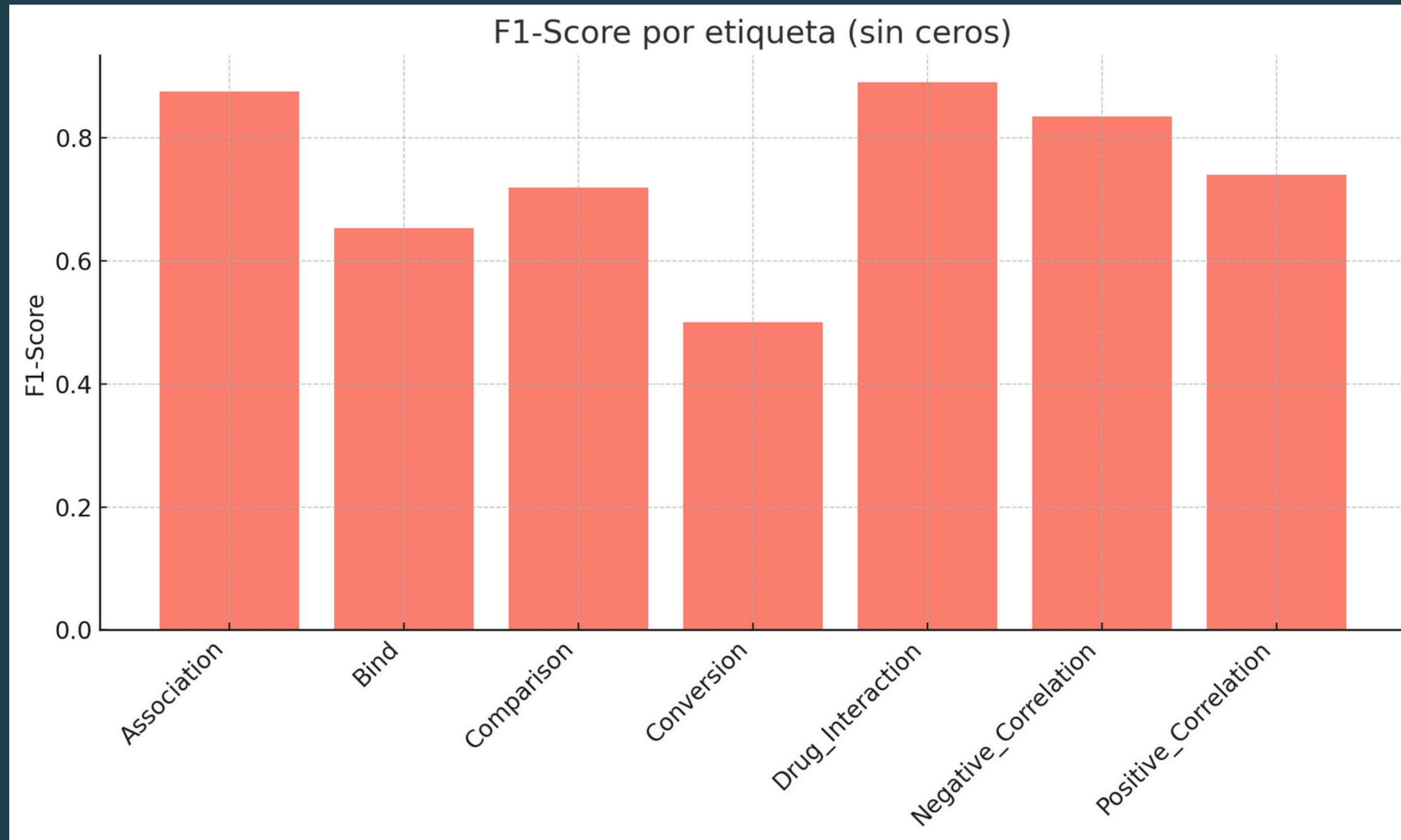
PRECISIÓN POR CATEGORÍA



RECALL POR CATEGORIA



F-1 POR CATEGORIA



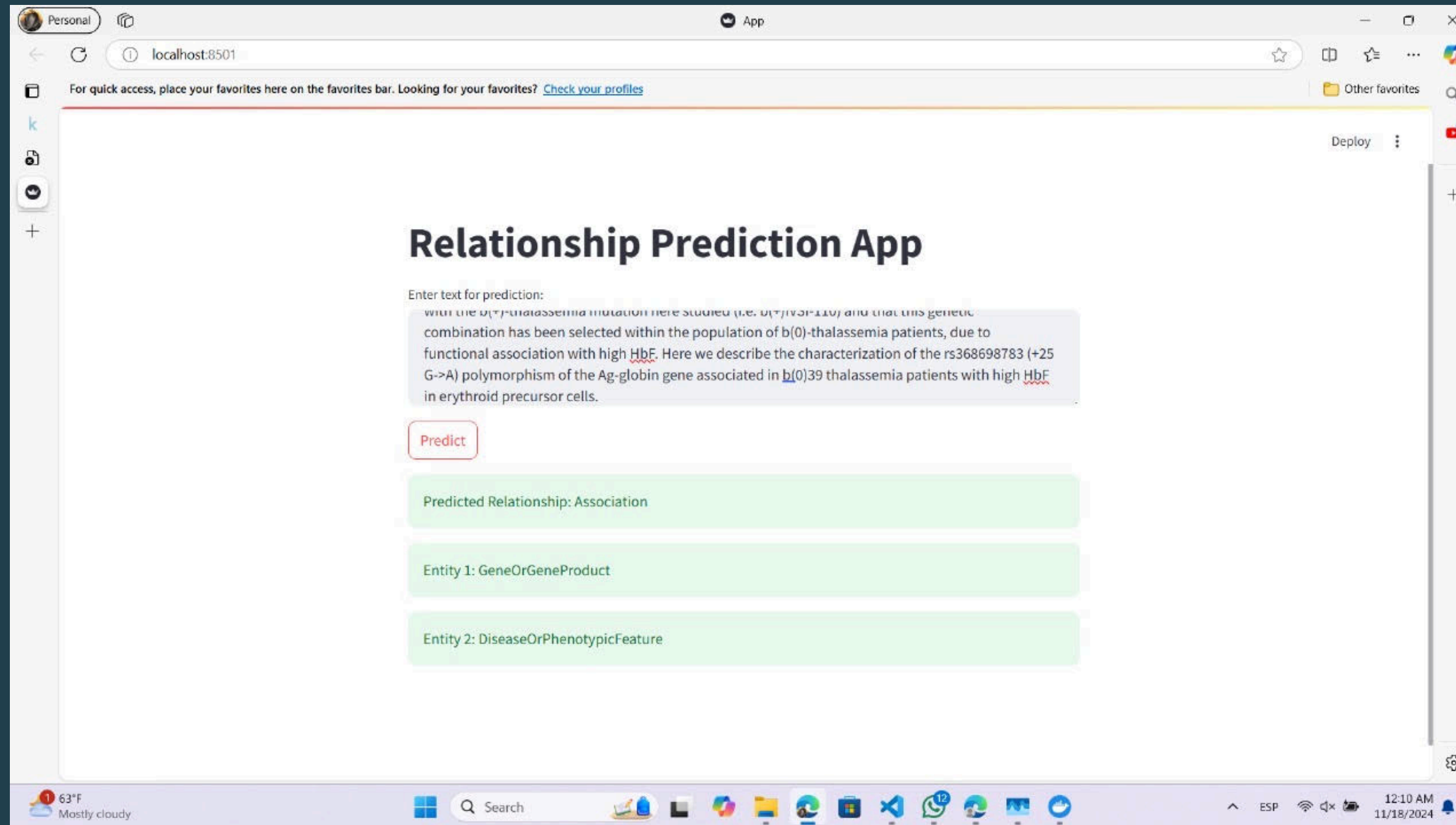
CONCLUSIONES

El modelo muestra buen desempeño en etiquetas comunes, pero falla en las menos representadas.

El desequilibrio de datos afecta negativamente las métricas globales.

Alcanza alta precisión (~85%) con estabilidad razonable en validación.

APLICACION WEB



GRACIAS