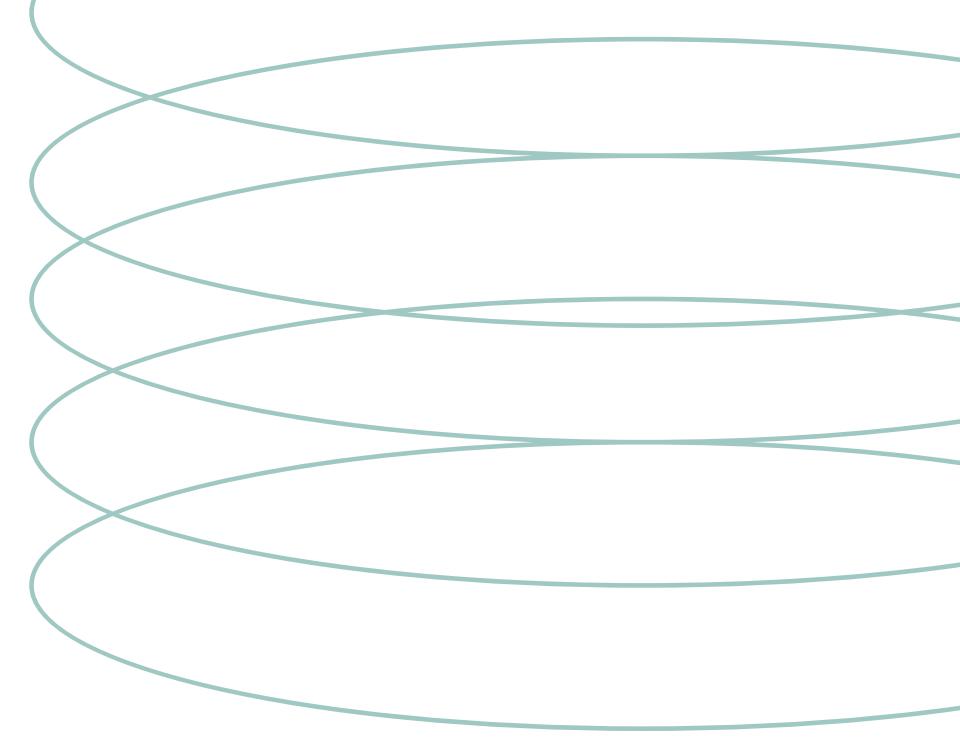


MAESTRÍA EN INTELIGENCIA ARTIFICIAL

Aprendizaje Automático II

Presentación de avances

Hernán Marano
Jorge Flores
Nicolás Velazquez
Rubén Ortigoza



Evaluación automática de ensayos científicos mediante predicción de calificaciones

En los congresos científicos, los resúmenes permiten difundir investigaciones de forma ágil. Su evaluación, clave para garantizar calidad y relevancia, depende de revisores expertos que trabajan en plazos acotados.

OBJETIVO

Determinar la viabilidad de aplicar métodos de aprendizaje automático para la evaluación de resúmenes científicos, utilizando un conjunto de entrenamiento que permita obtener métricas de calidad análogas a las de evaluaciones humanas.

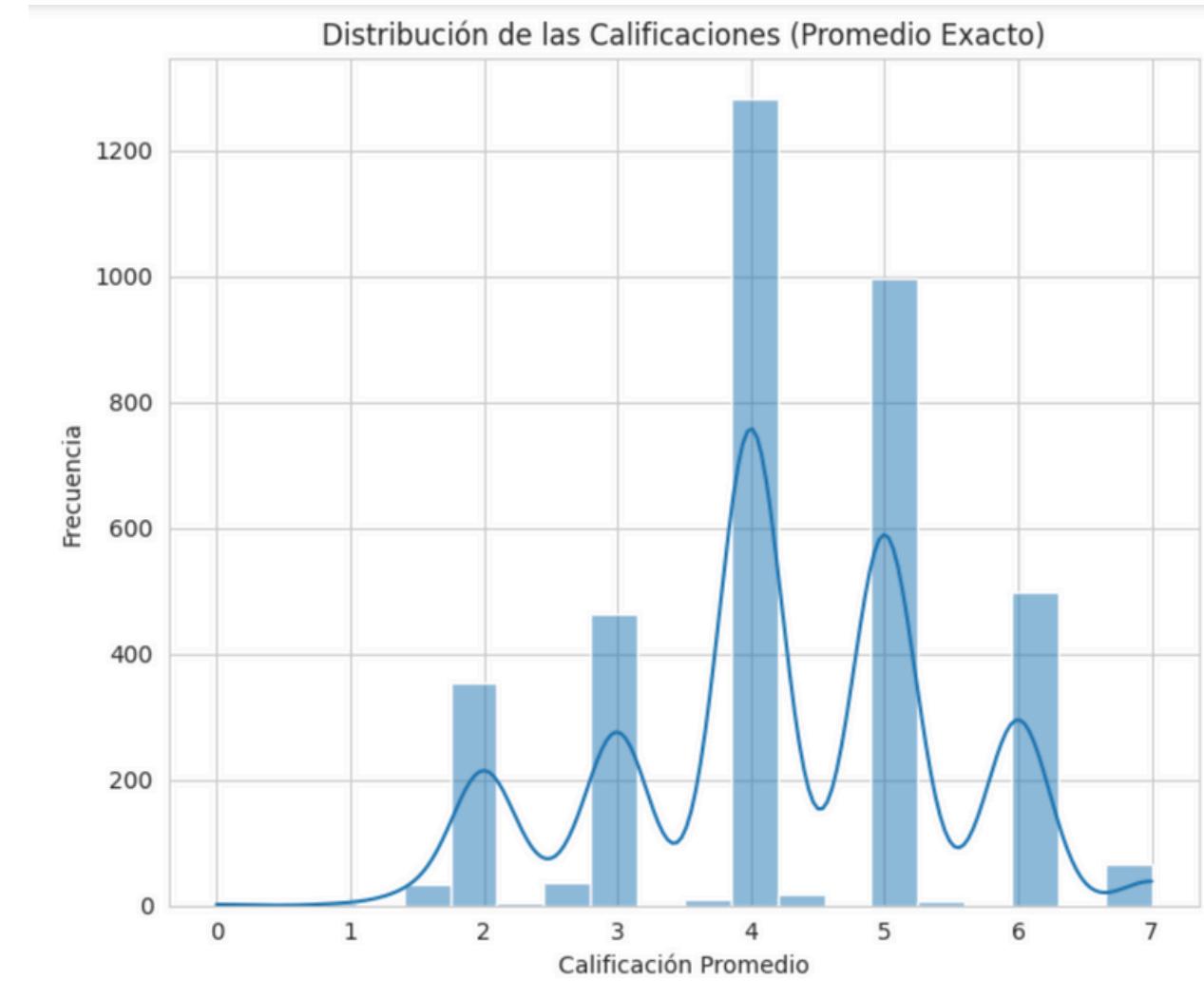
Dataset

Cantidad de Registros
3755

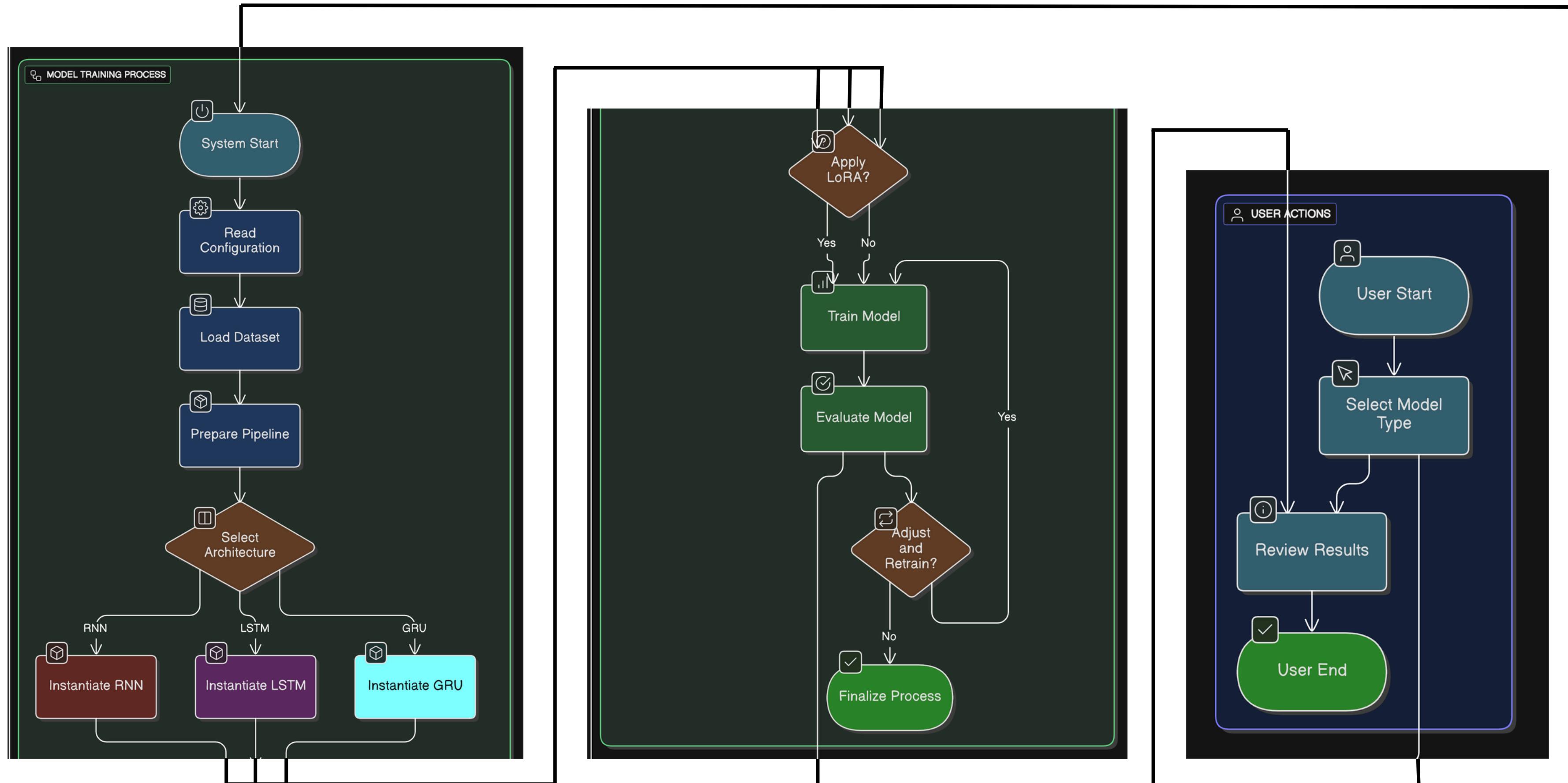
Features

- Título
- Autores
- Instituciones
- Resumen
- Área Temática
- Tipo de Presentación
- Puntaje Promedio

calificación	
count	3775.00
mean	4.22
std	1.22
min	0.00
25%	4.00
50%	4.00
75%	5.00
max	7.00



Flujo del proyecto

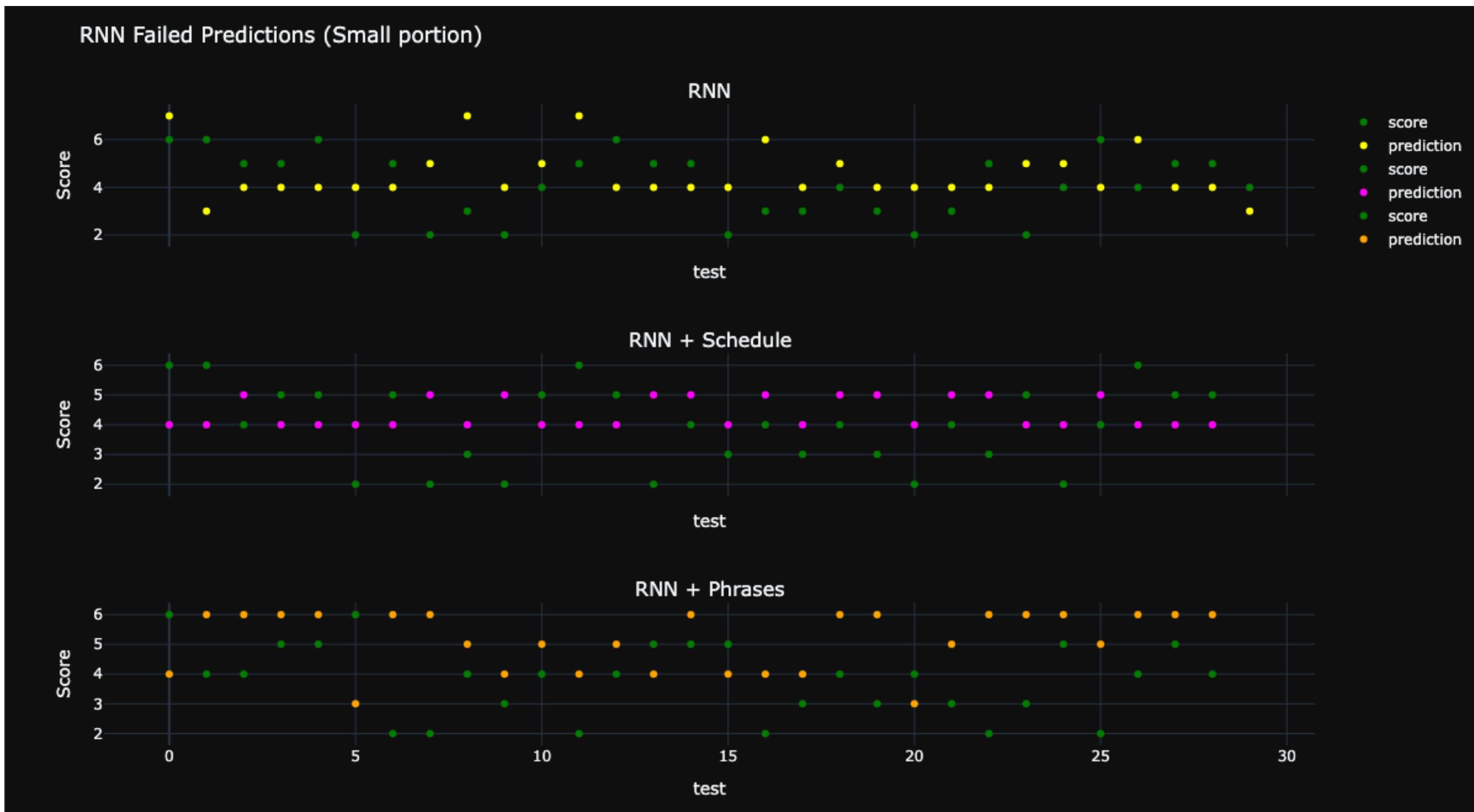


Rnn

modelo	epochs	lr	layer	hidden dim	embed dim	accuracy	f1 score	mse
RNN	200	0.001	1	128	100	24.61%	0.104	2.973
RNN+LR	200	0.001	2	128	100	17.54%	0.149	4.667
RNN+text	300	0.001	2	128	100	21.93%	0.124	4.333

[1]: Distributed Representation of words, phrases and their compositionality

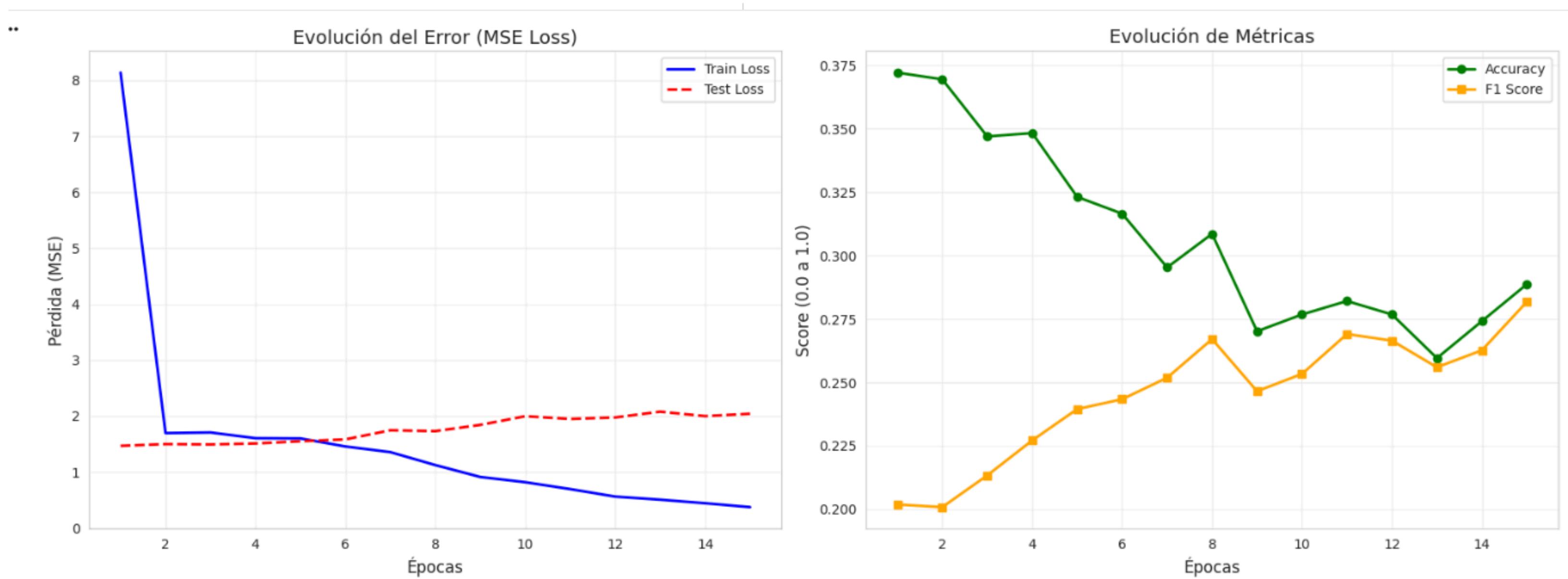
Grafico de predicciones Fallidas



LSTM

MODELO	EPOCHS	LR	LAYER	HIDDEN DIM	EMBED DIM	ACCURACY	F1 SCORE	MSE
Modelo 1	12	0.001	1	64	128	32.55%	0.187	2.612
Modelo 2	20	0.0005	2	128	128	30.12%	0.176	2.944
Modelo 3	8	0.001	1	96	128	28.47%	0.158	3.884
Modelo 4	10	0.001	1	64	128	26.91%	0.142	2.731

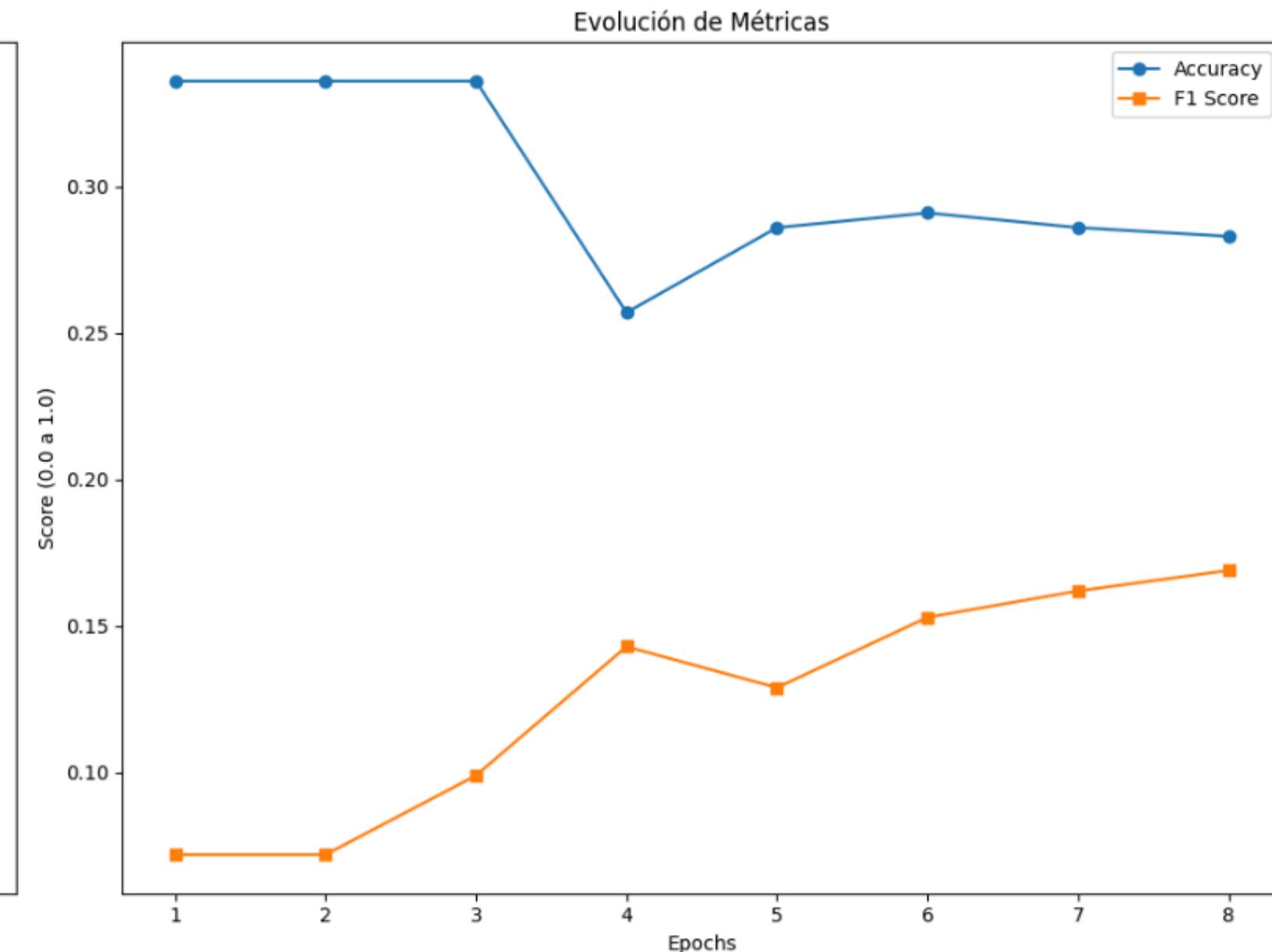
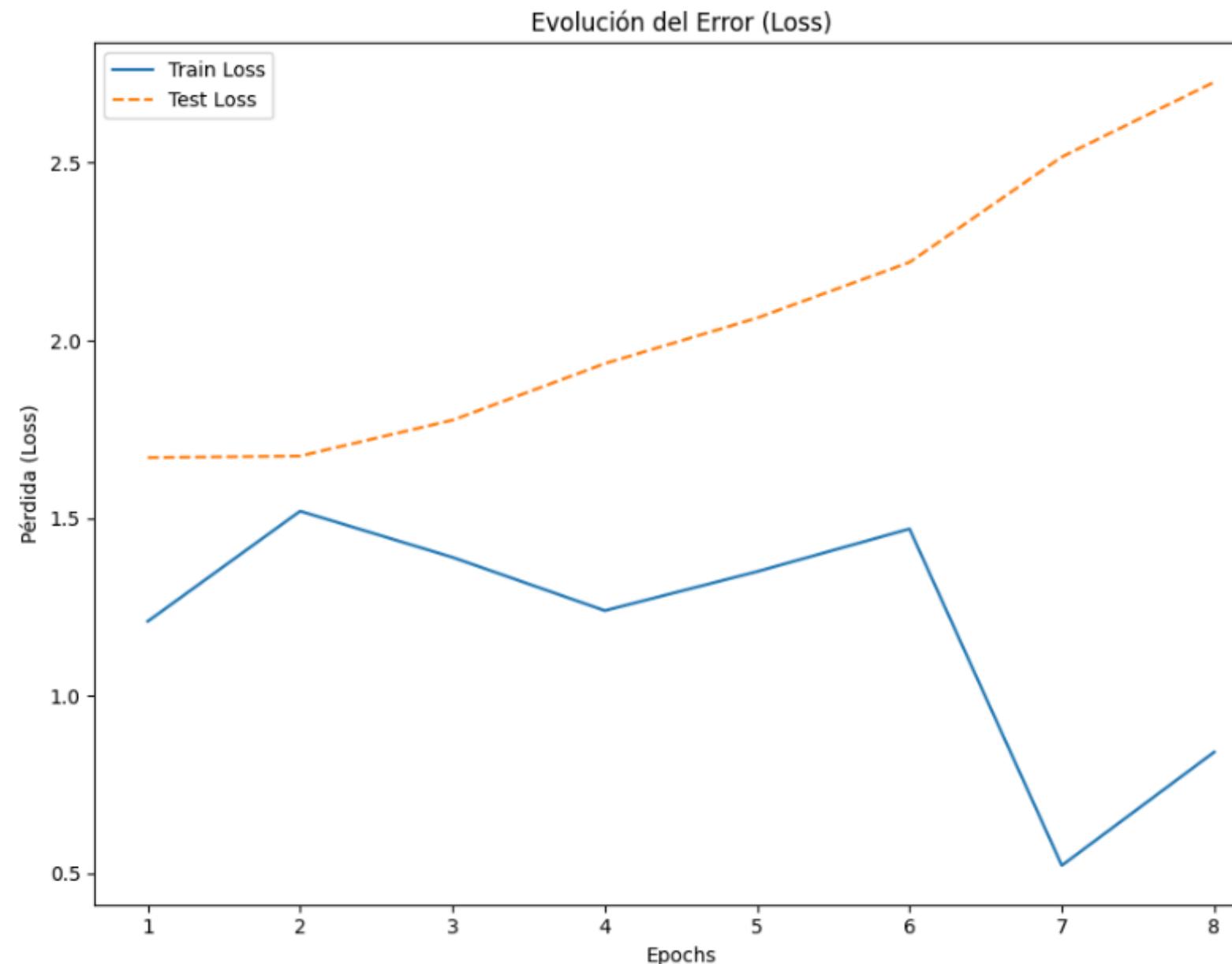
Grafico del modelo LSTM 1 (Best)



GRU

MODELO	EPOCHS	LR	LAYER	HIDDEN DIM	EMBED DIM	ACCURACY	F1 SCORE	MSE
Modelo 1	10	0.001	1	64	128	28.31%	0.169	2.7272
Modelo 2	20	0.0005	2	128	128	29.63%	0.147	2.8387
Modelo 3	8	0.001	1	96	128	25.40%	0.138	4.5446
Modelo 4	10	0.001	1	64	128	25.40%	0.131	2.4531

Grafico del modelo GRU



COMPARACIÓN DE RESULTADOS

MODELO (MEJOR VARIANTE)	ACCURACY	F1 SCORE	MSE	TIEMPO DE ENTRENAMIENTO (SEG)
RNN (con Scheduler)	24.61%	0.104	2.973	86
LSTM (variante 1)	32.55%	0.187	2.612	50
GRU (variante 2)	29.63%	0.147	2.8387	131
Transformers (BERT)	32.25%	0.183	2.35	1.8×10^4

Conclusiones

- Limitaciones de las RNN tradicionales
- Baja capacidad de generación
- Dependencia de patrones superficiales
- ECM elevado
- Necesidad de modelos más robustos

Gracias!