

 Analisis Exploratorio

RESULTADOS DEL LITCOIN NPL CHALLENGE

Analisis de datos del LitCoin NPL Challenge por
Jeyner Arango, Estefania Elvira, Sofia Escobar, Jose Monzon

Grupo 8



📶 Resultados Iniciales

ASPECTOS CLAVE



Deteccion de Conceptos cientificos

Se buscan sistemas que puedan reconocer con precision conceptos cientificos clave de textos cientificos biomedicos.



Identificacion de Relaciones Biomedicas

Identificar todas las relaciones entre el titulo y el resumen basadas en un modelo predictivo BioLink.



📶 Analisis Exploratorio

LIMPIEZA

- Se eliminaron los valores Nulos
- Se convirtieron los textos a minúsculas para mas uniformidad
- Se eliminaron caracteres especiales y puntuacion
-

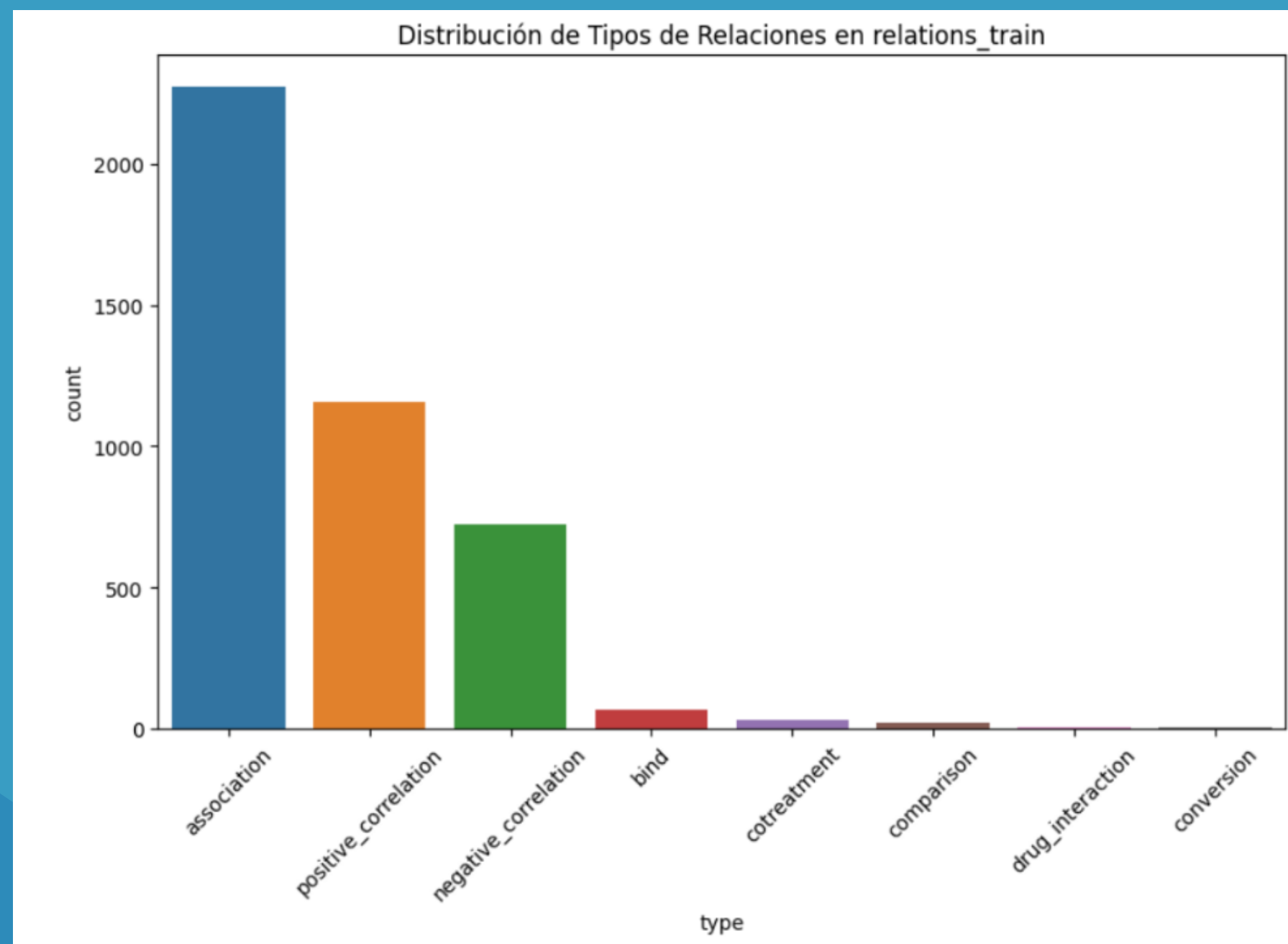


Analisis Exploratorio

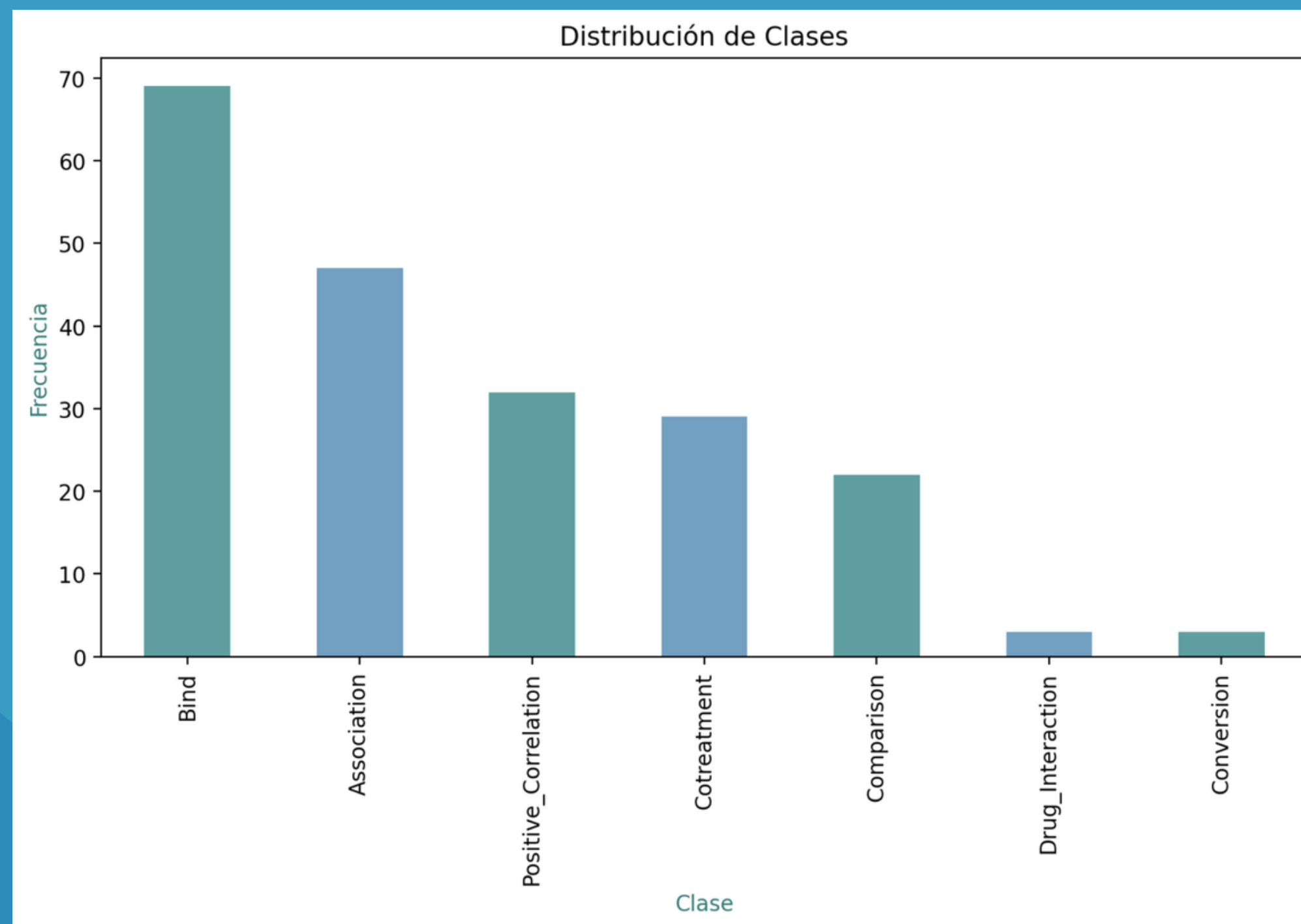


```
Tabla de frecuencia de tipos de relaciones:  
association          2274  
positive_correlation 1159  
negative_correlation  721  
bind                  69  
cotreatment           29  
comparison            22  
drug_interaction       3  
conversion             3  
Name: type, dtype: int64
```

Analisis Exploratorio



Ajustes en los datos



 resultados iniciales



ALGORITMOS

Los algoritmos que utilizamos fueron



Asignación de Dirichlet latente (LDA)

Es capaz de identificar temas latentes en los datos y asignar documentos a esos temas. Esto puede ser útil para tareas como la clasificación de documentos o la agrupación.



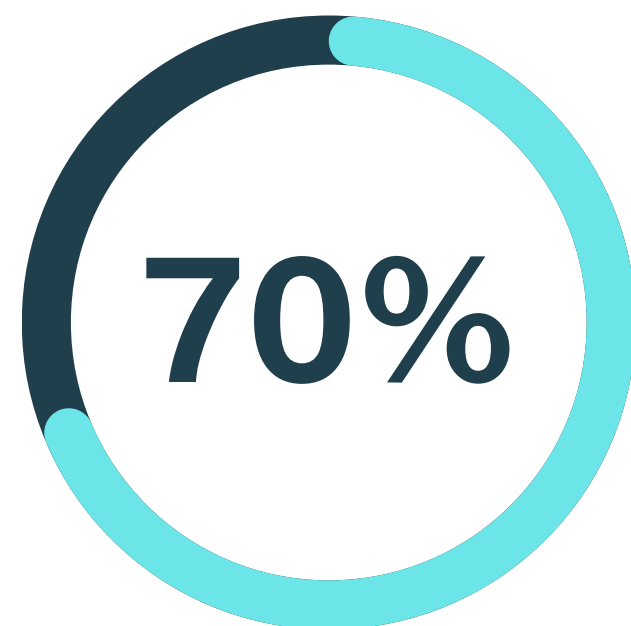
NLP

Este modelo de `en_core_web_sm` incluye funciones como tokenización, lematización, análisis morfológico, reconocimiento de entidades, y más.

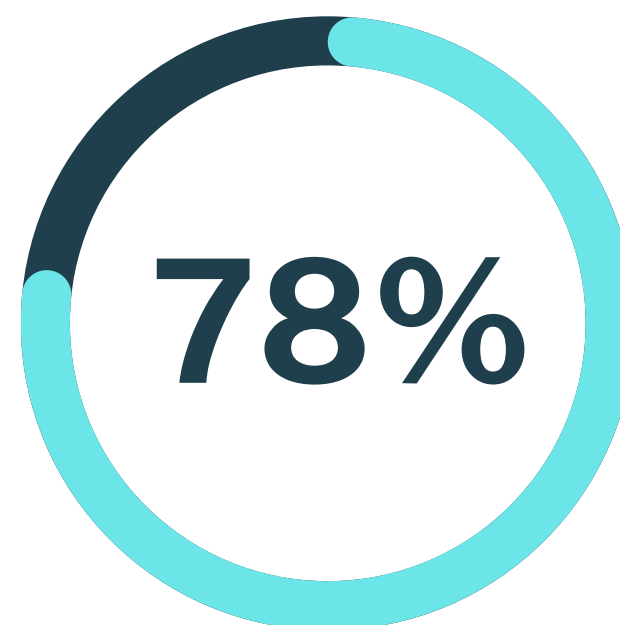
EFICACIA



Asignación de
Dirichlet latente (LDA)



NLP
en_core_web_sm
con SpaCY

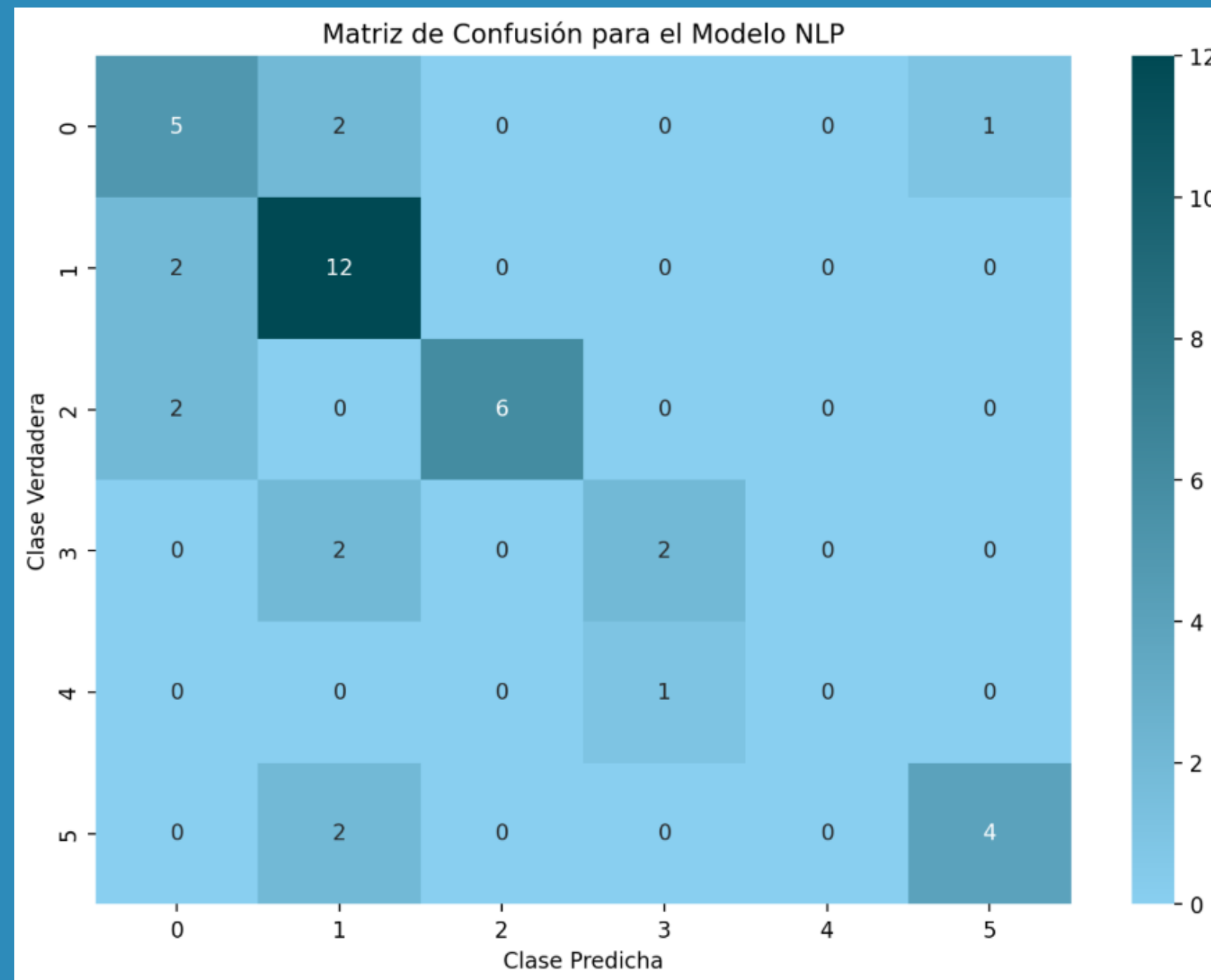




Resultados iniciales



MODELO NLP

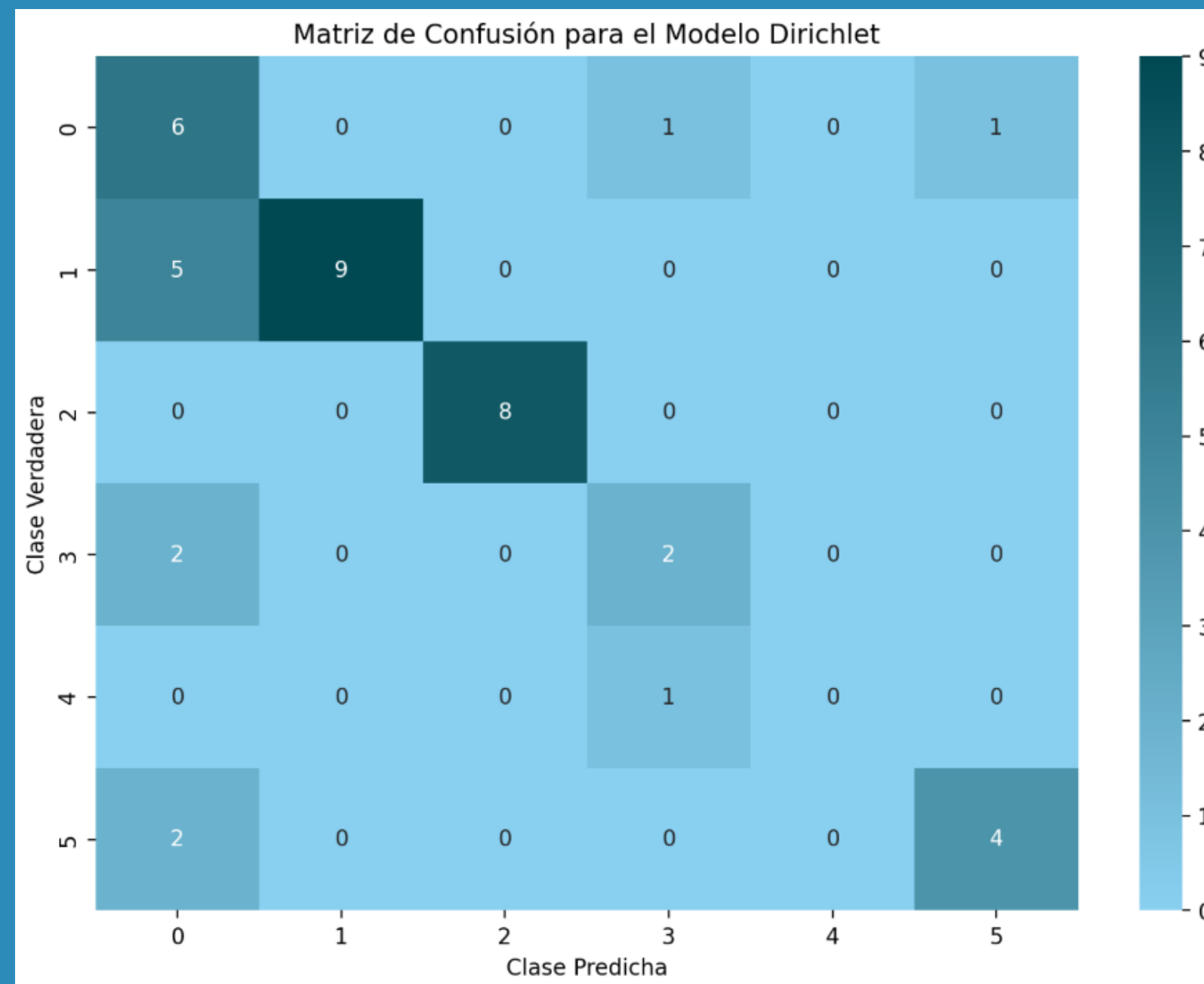




Resultados iniciales



MODELO LDA



📶 Analisis Exploratorio

CONCLUSIÓN

El proyecto cumple con la tarea de clasificación de relaciones en textos biomédicos. Los aspectos positivos destacados incluyen la implementación de herramientas efectivas de visualización y la interactividad integrada a través de Streamlit, lo que contribuye significativamente a la comprensión y presentación de los resultados obtenidos.





Resultados iniciales

GRACIAS +



Grupo 8