Analisis Exploratorio

ANALISIS DE DATOS DEL LITCOIN NPL CHALLENGE

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

Grupo 8

Analisis Exploratorio

SITUACION PROBLEMATICA

Es necesario acelerar la investigfacion cientifica en el area de la medicina mediante el uso de la ciencia de los datos.

Analisis Exploratorio

ASPECTOS CLAVE

Deteccion de Conceptos cientificos

Se buscan sistemas que puedan reconocer ocn precision cconceptos 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

PROBLEMA CIENTIFICO

El problema cientifico consta en desarrollar sistemas de procesamiento de lenguaje natural avanzados que puedan comprender y analizar textos cientificos de manera eficiente para extraer informacion, identificar relaciones y determinar la novedad de los hallazgos.

Analisis Exploratorio

OBJETIVOS

Nuestros principales objetivos para la elaboracion del proyecto

Desarrollar modelos NLP

Desarrollar modelos de procesamiendo del lenguaje natural, que sean capaces de procesar textos cientificos.

Determinar novedad cientifica

Elaborar sistemas capaces de avaluar si las relaciones identificadas presentan informacion reelevante

Reconocimiento de Entidades

Se quiere lograr que los sistemas puedan reconocer correctamente las entidades biomedicas presentes

Analisis Exploratorio

Entidades (13,600 datos)

`id`, `abstract_id`, `offset_start` y `offset_finish`, `type`, `mention`: `entity_ids`

Relaciones (2,300 datos)

- `id`, `abstract_id`, `type`, `entity_1_id` y `entity_2_id`, `novel`

Abstract (400 datos)

- `abstract_id`, `title`,`abstract`

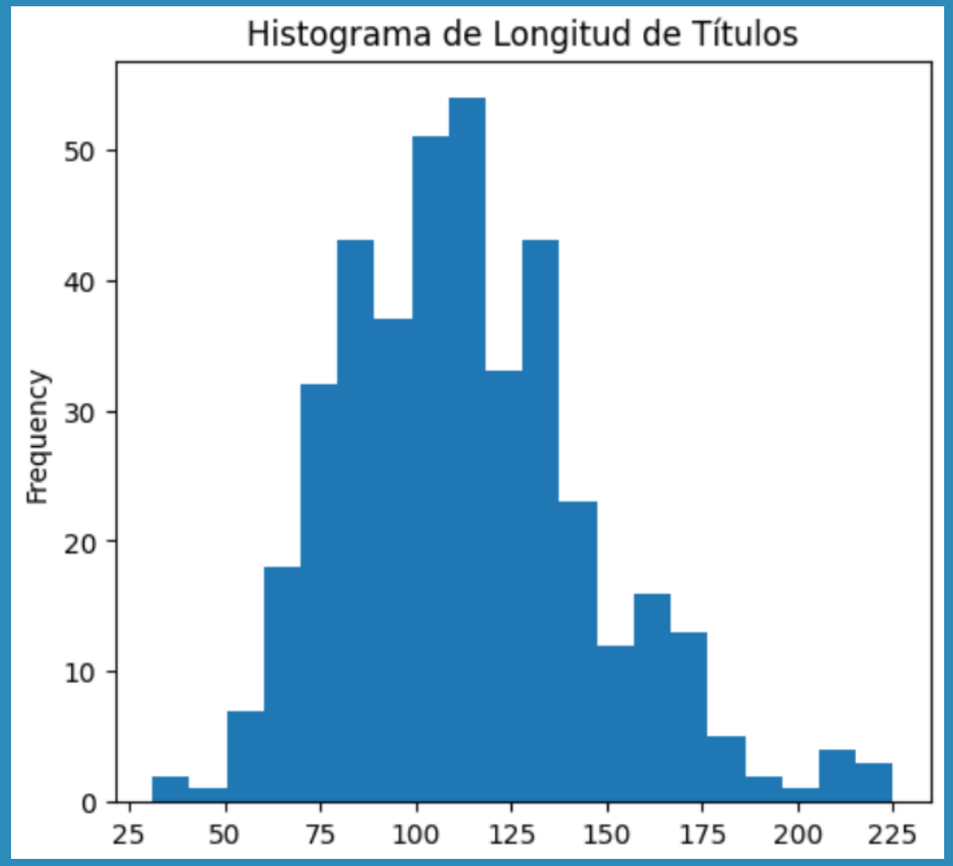
Analisis Exploratorio

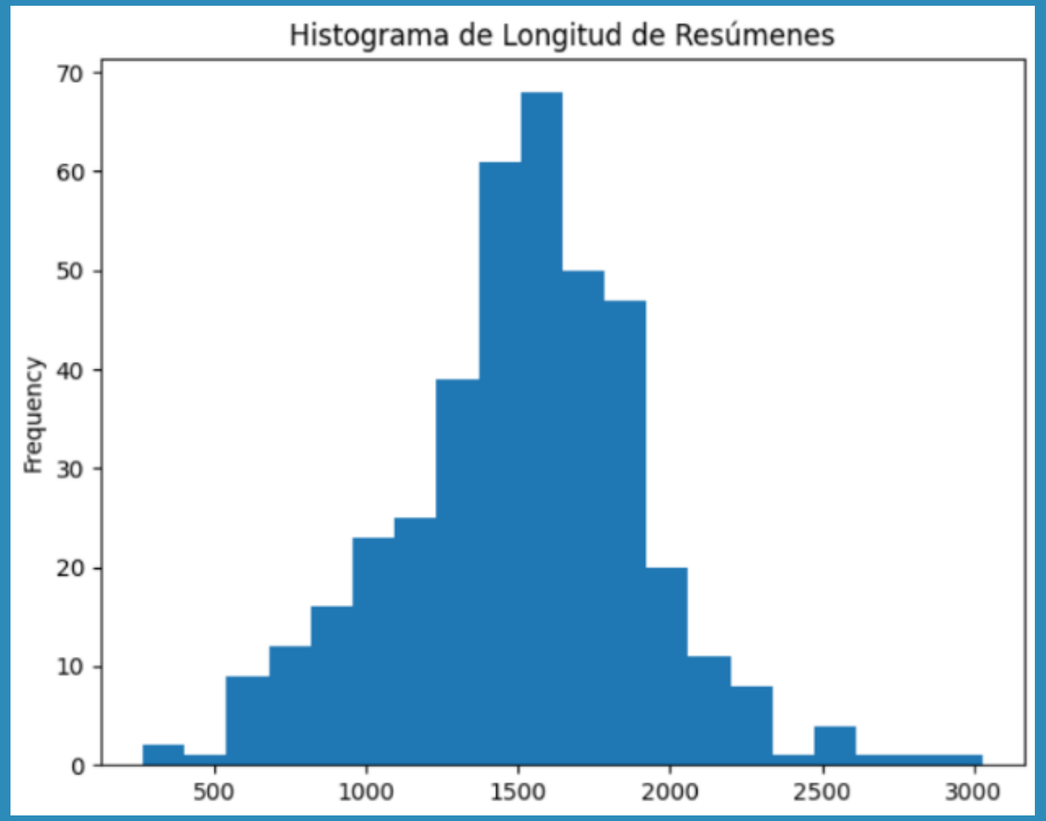
LIMPIEZA

- Se eliminaron los valores Nulos
- Se convirtieron los textos a minusculas para mas uniformidad
- Se eliminaron caracteres especiales y puntuacion
- Se eliminaron las stopwords utilizando la base de datos de ntl.corpus

Analisis Exploratorio

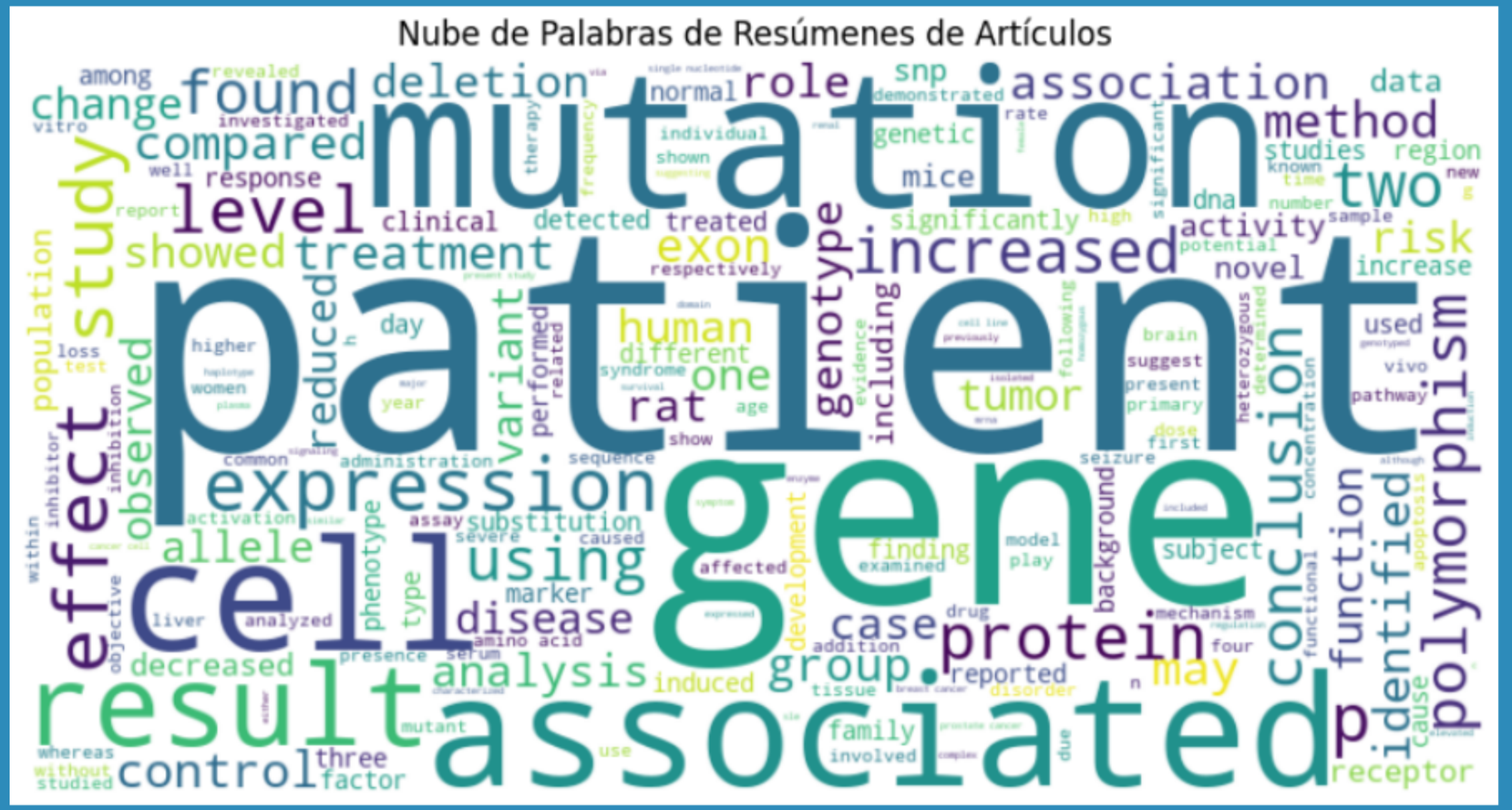
ABSTRACT





Analisis Exploratorio

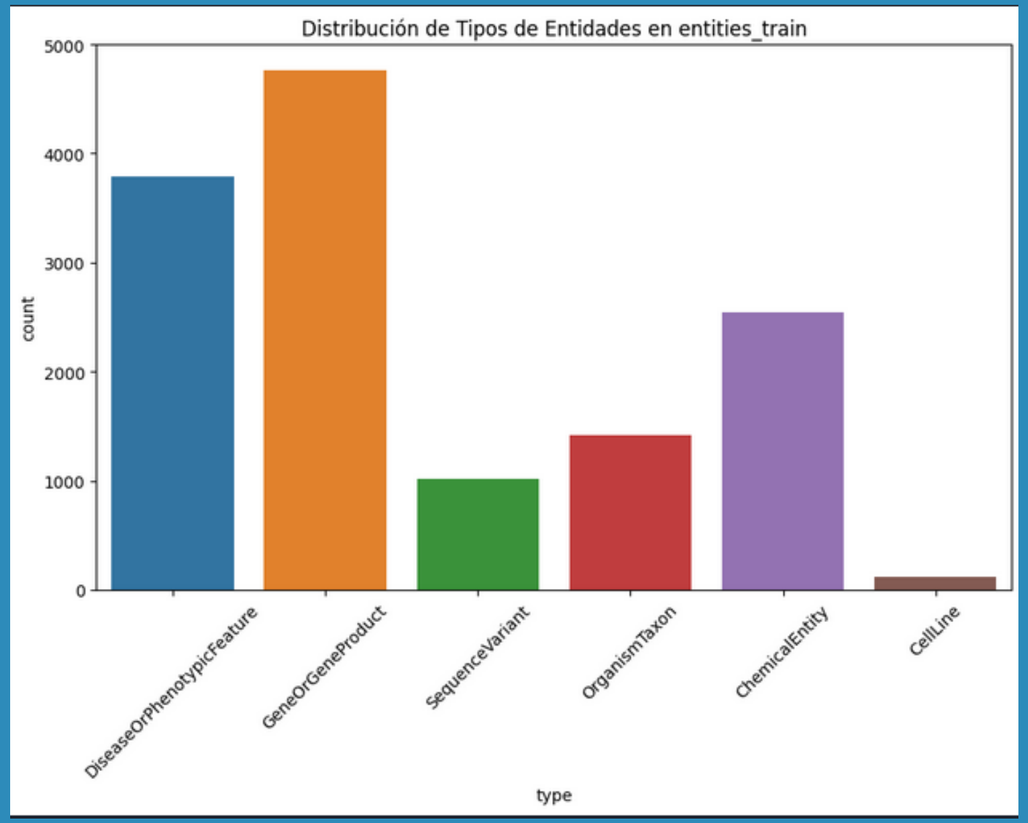
ABSTRACT



Analisis Exploratorio

ENTIDADES

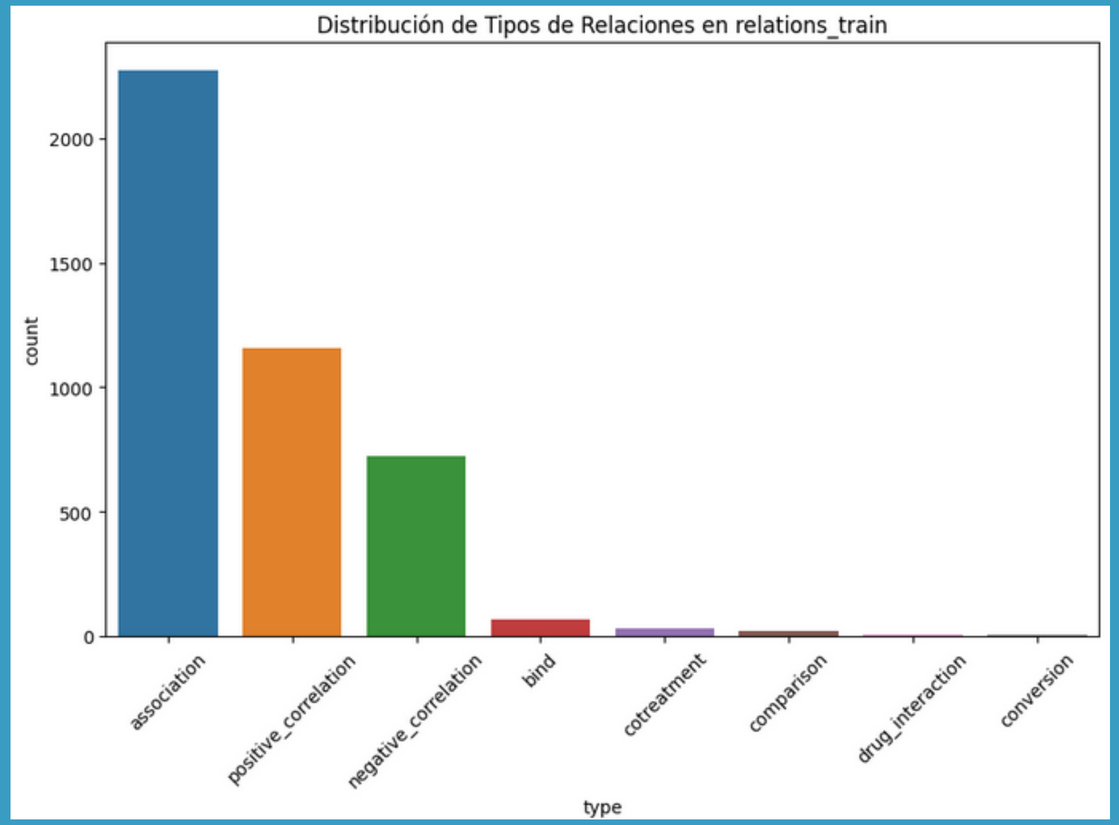
```
Tabla de frecuencia de tipos de entidades:
GeneOrGeneProduct             4764
DiseaseOrPhenotypicFeature    3784
ChemicalEntity                2540
OrganismTaxon                 1420
SequenceVariant               1011
CellLine                       117
Name: type, dtype: int64
```



Analisis Exploratorio

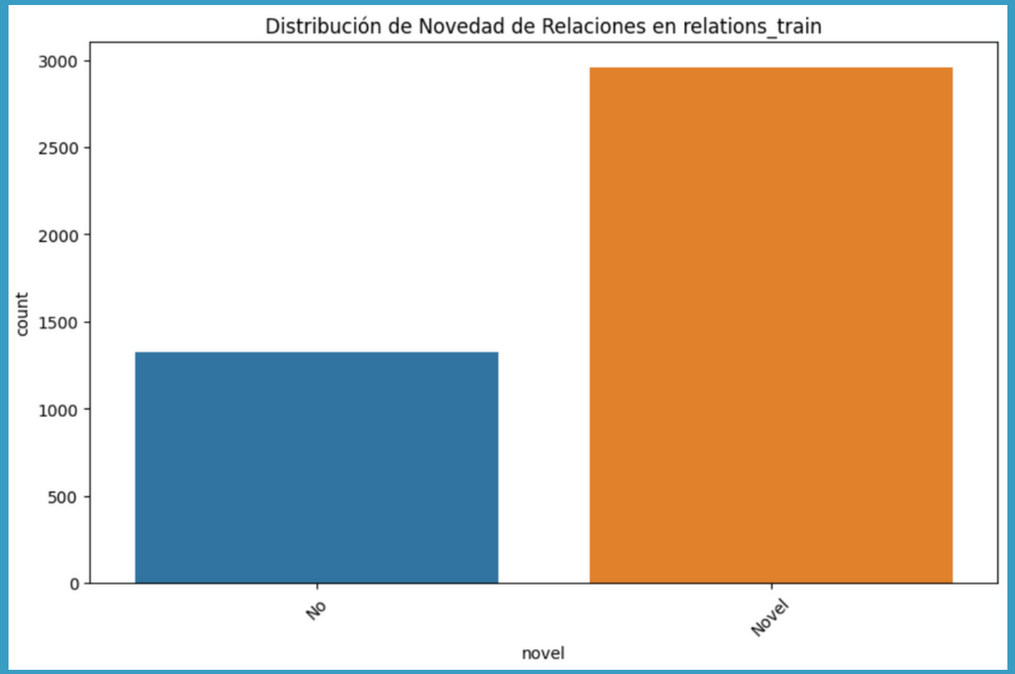
RELACIONES

```
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

RELACIONES



Analisis Exploratorio

CONCLUSIONES

- Los datos de entidades y relaciones proporcionan información importante, lo que es esencial para el problema planteado.

- Los datos de texto contienen una variedad de temas y requieren tecnicas avanzadas de NLP.

- Se podría considerar la implementación clasificación de texto o la extracción de información, para abordar el desafío de identificar relaciones en los textos de los artículos.

Analisis Exploratorio

GRACIAS

Grupo 8