



Universidad del Desarrollo
Facultad de Ingeniería

POPULISMO Y MIGRACIÓN: ANÁLISIS DE REDES DE DISCURSOS
PRESIDENCIALES MEDIANTE PROCESAMIENTO DE LENGUAJE NATURAL

POR: DANIEL IGNACIO MORALES ROJAS

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dra. MARIA PAZ RAVEAU MORALES

Diciembre 2024

SANTIAGO

A mi madre, Lorena, y a mi padre, Patricio, por la oportunidad y el apoyo que me han brindado para alcanzar este logro.

A mi pareja, Jossefa, por su apoyo incondicional en los momentos difíciles y por ser mi mayor fortaleza.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	2
2. TRABAJO RELACIONADO	4
3. HIPÓTESIS Y OBJETIVOS	6
3.1. HIPÓTESIS	6
3.2. OBJETIVOS	6
<i>Objetivo general.....</i>	<i>6</i>
<i>Objetivos específicos</i>	<i>6</i>
4. DATOS Y METODOLOGÍA	8
4.1. DATOS	8
<i>Base de Datos Global Populism Database v2.....</i>	<i>8</i>
<i>Discursos</i>	<i>11</i>
4.2. METODOLOGÍA	11
<i>Preprocesamiento de datos</i>	<i>14</i>
<i>Análisis de texto</i>	<i>15</i>
<i>Creación de redes.....</i>	<i>16</i>
<i>Creación de comunidades</i>	<i>17</i>
<i>Predicción de nivel de populismo</i>	<i>17</i>
5. RESULTADOS.....	19
5.1. PREPROCESAMIENTO DE DATOS.....	19
5.2. RESULTADOS DEL ANÁLISIS DE TEXTO	20
5.3. CONSTRUCCIÓN DE RED	21
5.4. GENERACIÓN DE COMUNIDADES.....	22

5.5.	PREDICCIÓN DE NIVEL DE POPULISMO	24
5.6.	ANÁLISIS COMPLEMENTARIO	27
6.	CONCLUSIONES	29
6.1.	PROPUESTAS PARA TRABAJOS FUTUROS	30
	BIBLIOGRAFÍA	31

Resumen

El presente trabajo explora la conexión entre el populismo y las referencias a asuntos migratorios en discursos de gobiernos de América Latina, empleando técnicas de Procesamiento de Lenguaje Natural (NLP), análisis de redes y métodos de Machine Learning. El estudio examina un corpus de 273 discursos pronunciados entre 1998 y 2019. Mediante el estudio de textos y la detección de patrones discursivos, se crearon redes que representan menciones entre naciones, valorando métricas fundamentales como la centralidad y la frecuencia de términos vinculados a la migración.

Se crearon agrupaciones de países basadas en parámetros políticos y métricas de centralidad calculadas durante el armado de las redes. Además, se creó un modelo predictivo fundamentado en Gradient Boosting, que tiene la capacidad de calcular los niveles de populismo con una precisión del 63%. Los hallazgos indican que los discursos que abordan asuntos migratorios suelen mostrar ciertas tendencias de populismo, lo cual puede inferir que estas narrativas se utilizan de manera estratégica para captar respaldo político (Martínez, 2023).

Así, este análisis aporta a la comprensión del uso de narrativas migratorias en escenarios políticos, marcando la influencia en la formación de discursos populistas.

1. Introducción

Los temas de migración en el continente americano se han convertido en un tema cada vez más controversial. Según datos recientes, la migración en países de América Latina ha aumentado considerablemente desde finales de los 90', con alrededor de 11 millones de migrantes en América Latina originarios de otros países dentro de la misma región (McAuliffe y Oucho, 2024).

Este fenómeno ha adquirido más importancia en la actualidad, en un escenario en el que los movimientos populistas están en aumento a escala mundial (Correia, 2023) y las dinámicas migratorias se ven cada vez más. El populismo se define, desde una perspectiva política, como un modo de construir lo político mediante la articulación de demandas populares, donde el líder representa al "pueblo" en oposición a las élites, pudiendo tener un carácter transformador o excluyente (Martinez A., 2023). En este sentido, el contexto migratorio sirve como punto de partida para analizar cómo los discursos presidenciales en América Latina abordan las narrativas migratorias y cómo estas menciones pueden influir en los niveles de populismo ya que la migración es un tema polarizador, que permite construir a líderes narrativas de 'nosotros contra ellos', algo central dentro de los discursos populistas (Martínez, 2023).

Así, el estudio de esta se basa en la información obtenida de la Base de Datos de Populismo Global v2, la cual recolecta discursos de 241 líderes de 74 naciones diferentes, codificados con una calificación de populismo mediante métodos de evaluación holística (Hawkins et al, 2019). La importancia del estudio radica en su habilidad para proporcionar

un entendimiento detallado de cómo los asuntos migratorios se incorporan en las narraciones populistas, y de esta manera verificar la hipótesis de que las referencias concretas a la migración suelen representar niveles más elevados de populismo en los discursos del presidente.

Para analizar esta relación, se emplearán herramientas avanzadas de procesamiento de lenguaje natural (NLP), Machine Learning (ML) y análisis de redes. Estas metodologías permitirán identificar patrones discursivos y evaluar cómo las narrativas relacionadas con la migración pueden estar asociadas a discursos más populistas.

2. Trabajo Relacionado

Los temas principales que aborda este estudio son la temática del populismo, analizado a partir de la base de datos proporcionada y al tema de la migración, que fue seleccionado para explorar su presencia en los discursos entregados y para encontrar la relación posible que se tiene.

En este contexto, para cada tema, existen investigaciones o artículos previas que utilizan ciencia de datos para estudiar sobre los mismos temas con enfoques propios, como:

- a) “Discursos políticos y mediáticos contemporáneos sobre los inmigrantes”¹

El artículo si bien no tiene un enfoque sobre la ciencia de datos, si hace un análisis mediante lectura y revisión de publicaciones en temas de inmigración, donde en sus resultados se encontraron asociación de palabras de connotación negativa cuando los medios hablaban de inmigración, sobre todo en ámbitos laborales y económicos y como se ven afectados (Stuardo et al, 2021).

- b) “Identifying Populist Paragraphs in Text: A Machine-Learning Approach”²

El artículo presenta el de un modelo avanzado de clasificación mediante el uso de BERT (el NLP de Google) que es capaz de identificar contenido populista en textos políticos, permitiendo al desarrollo de herramientas automatizadas para el análisis de contenido. Se indica como conclusión que, aunque el modelo no es completamente autónomo, puede ayudar a los codificadores humanos mediante un sistema "humano en el bucle", donde el

¹ Disponible en <https://revistas.comillas.edu/index.php/revistamigraciones/article/view/12935>

² Disponible en <https://arxiv.org/abs/2106.03161>

modelo entrega párrafos para que el codificador pueda verificar, reduciendo significativamente el trabajo por parte de un ser humano (Ulinskaitė y Pukelis, 2021).

- c) “Estimación de los flujos migratorios mundiales mediante la combinación de fuentes de datos tradicionales e innovadoras y el aprendizaje automático”

Si bien este último estudio se relaciona más para posibles trabajos futuros, donde se pueda cruzar la información de los flujos migratorios para buscar correlaciones, es bueno incluir que existe la información en forma de base de datos y cuantificable en el Portal de Datos Sobre Migración³

³ Disponible en <https://www.migrationdataportal.org/es/data-innovation-38>

3. Hipótesis y Objetivos

3.1. Hipótesis

Los discursos que incluyen menciones explícitas de temas migratorios presentan una mayor probabilidad de tener altos puntajes de populismo. Esto se estudiará a través de revisión en la frecuencia de dichas menciones dentro del corpus de discursos políticos de la Global Populism Database.

3.2. Objetivos

Objetivo general

Explorar mediante técnicas de procesamiento de lenguaje natural (NLP), análisis de redes, creación de comunidades y machine learning, la relación entre los puntajes de populismo en discursos presidenciales y las menciones a temas migratorios extraídos de los discursos de la Global Populism Database. Este análisis buscará identificar conexiones narrativas y comunidades para poder evaluar las posibles relaciones entre narrativas populistas y los temas migratorios.

Objetivos específicos

- Usar técnicas de Named Entity Recognition (NER) para identificar menciones a países y temas relacionados con migración en los discursos.
- Construir una red dirigida que relacione las menciones entre países para identificar relaciones discursivas y países más influyentes.

- Aplicar algoritmos de detección de comunidades para identificar grupos con narrativas similares, explorando patrones a través de comunidades de países densamente conectados en la red discursiva.
- Identificar las principales variables predictoras y desarrollar un modelo Machine Learning que permita predecir puntaje populismo de los discursos.

4. Datos y Metodología

4.1. Datos

El estudio utiliza los datos proporcionados por la Global Populism Database v2 ⁴, el cual incluye una base de datos que asigna niveles de populismo a discursos de líderes políticos, la colección de discursos en formato de texto completo. A esto se le agrega un diccionario especializado llamado Moral Foundation (Haidt y Graham, 2007) en español que contiene palabras que pueden identificar temas relacionados con migración. La combinación de estos recursos posibilita un análisis integral de las narrativas políticas y su relación con el populismo, así como un enfoque particular en los discursos relacionados con temas migratorios.

Base de Datos Global Populism Database v2

Esta base de datos contiene la información general de los discursos (Tabla 1) con la asignación de nivel de populismo de cada discurso que ha sido codificado manualmente por 1 o más especialistas, utilizando técnica de clasificación holística, garantizando un análisis estandarizado de los niveles de populismo en el discurso político (Hawkins et al, 2019).

⁴ [Data set] Global Populism Database (Versión 2.0). Harvard Dataverse. Disponible en: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/LFTQEZ>

Tabla 1. Descripción de las variables en la base de datos

Nombre variable	Descripción	Nombre variable en español
merging_variable	Nombre del archivo .txt	nombre_archivo
Country	País	País
Leader	Nombre del presidente o primer ministro	líder
Party	Partido o grupo político al cual está afiliado el líder	Partido
lr	Indicador categórico de orientación política, donde -1 = izquierda, 0 = centro, 1 = derecha	espectro_politico
president	Indicador binario: 1 si el líder es presidente y 0 si es primer ministro	tipo_lider
term	Periodo de mandato del líder en el cargo	periodo
startofterm	Fecha de inicio del periodo de mandato	inicio_periodo
yearbegin	Año de inicio del mandato	año_inicio
endofterm	Fecha de fin del periodo de mandato	fin_periodo
yearend	Año de fin del mandato.	año_fin
speechtype	Tipo de discurso: international, campaign, ribbon-cutting, famous	tipo_discurso
speechnum	Número de discurso para el mandato del líder	codigo_tipo_discurso
codernum	Identificador del codificador	codificador
rubricgrade	Nota del discurso por el codificador en la escala de populismo: (0 - 2)	nota_populismo
averagerubric	Promedio de todas las notas de los codificadores para ese discurso	nota_promedio_discurso
totalaverage	Promedio de todos los discursos de todos los codificadores para ese mandato del líder	nota_promedio_periodo
wb_region	Clasificación de región según el Banco Mundial (East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia, Sub-Saharan Africa)	region_wb
region	Subdivisión de regiones globales	región

La base de datos incluye 4.960 registros (Fig. 1), que contienen la información de 1240 discursos de 324 mandatos gubernamentales, 214 líderes y 74 países (Tabla 2), con discursos desde 1934 hasta el 2021 (Fig. 2). De estos nos centraremos netamente en los países de habla hispana de América Latina.

Distribución Global de Discursos por País

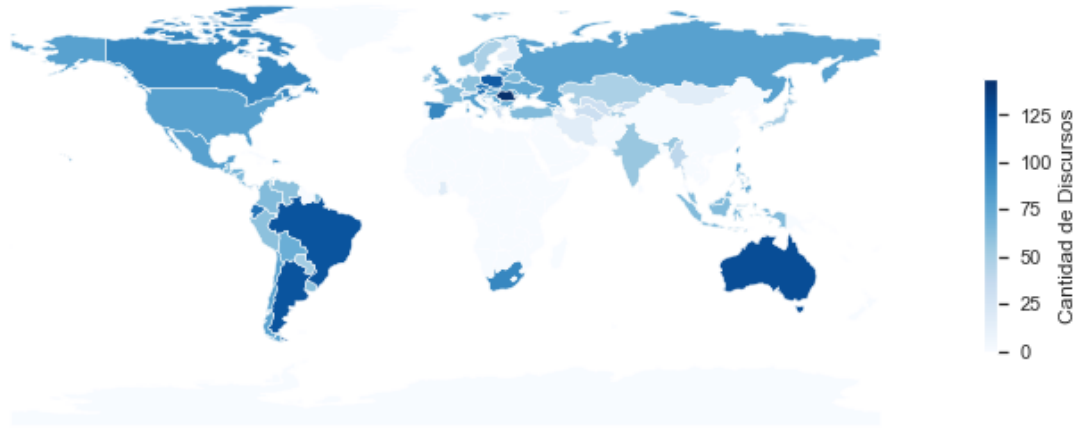


Fig. 1: Distribución de discursos por país

Tabla 2: Países de cada región

Región	Pais	Región	Pais
Europe & Central Asia	Albania	Latin America & Caribbean	Argentina
	Armenia		Bolivia
	Austria		Brazil
	Azerbaijan		Chile
	Belarus		Colombia
	Bulgaria		Costa Rica
	Croatia		Dominican Republic
	Czechia		Ecuador
	Estonia		El Salvador
	Finland		Guatemala
	France		Honduras
	Georgia		Mexico
	Germany		Nicaragua
	Greece		Panama
	Hungary	Paraguay	
	Ireland	Peru	
	Italy	Uruguay	
	Kazakhstan	Venezuela	
	Kyrgyzstan	East Asia & Pacific	Australia
	Latvia		Indonesia
	Lithuania		Japan
	Moldova		Malaysia
	Montenegro		Mongolia
	Netherlands		Myanmar
	North Macedonia		Philippines
	Norway	Singapore	
	Poland	Taiwan	
	Romania	North America	Canada
	Russia		United States
	Serbia	Middle East & North Africa	Iran
	Slovakia	South Asia	India
	Slovenia	Sub-Saharan Africa	Ghana
Spain	Africa	South Africa	
Sweden			
Switzerland			
Tajikistan			
Turkey			
Turkmenistan			
UK			
Ukraine			
Uzbekistan			

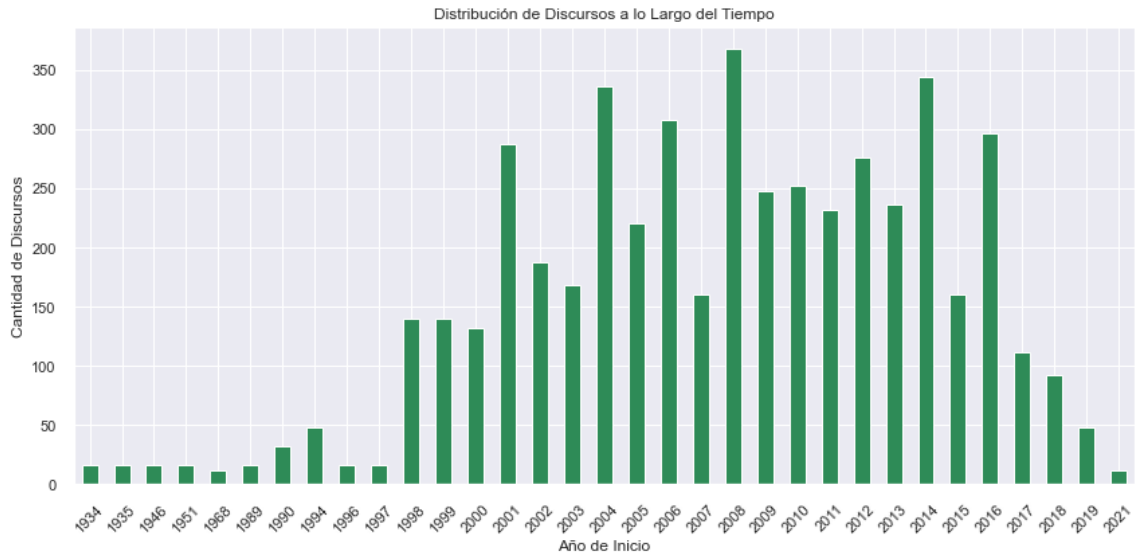


Fig. 2: Distribución de discursos en el tiempo

Discursos

El repertorio entrega 1.161 discursos en formato de texto. Los otros 79 discursos faltante son textos que no se encuentran disponibles por razones técnicas, por ejemplo, discursos evaluados a partir de vídeo o audio para los cuales no hay transcripción (Hawkins et al, 2019).

4.2. Metodología

El proyecto está basado con un enfoque en el análisis de textos. Así, para poder cumplir con los objetivos específicos propuestos, se realiza el pre procesamiento de la información para poder obtener información de calidad, se utiliza técnicas para identificación de menciones de países y palabras relacionadas con migración, en conjunto con diccionarios definidos para el análisis de los textos, construcción de redes dirigidas para poder analizar las conexiones entre países, división por clústeres para detectar comunidades con

narrativas similares y finalmente evaluar la relación de las variables para predecir el nivel de populismo en base a las variables más influyentes.

Las herramientas, modelos y técnicas para poder dar cumplimiento con los objetivos planteados fueron lo siguiente:

a) Diccionarios y Listas

- **Diccionario de Moral Foundation:** La teoría de Moral Foundation define cinco fundamentos morales, Care, Fairness, Ingroup, Authority y Purity (Haidt y Graham, 2007). Para este análisis, utilizamos palabras clave relacionadas con “Ingroup”, que hace referencia a compromiso grupal y con valores llamados “vice”, la cual capta referencias a términos clave relacionados con migración (ej.: "emigrar", "deserción", "migrar").
- **Lista de países y gentilicios:** Incluye nombres de países y gentilicios para detectar menciones geopolíticas. Se elabora a partir de los países que se incluyen en el Global Populism Database v2.

b) Reconocimiento de Entidades Nombradas (NER)

El Named Entity Recognition (NER) es una técnica del Procesamiento de Lenguaje Natural (NLP) que identifica y clasifica entidades en un texto, como nombres de personas, organizaciones, lugares, fechas o cantidades. Spacy proporciona modelos preentrenados para NER con alta precisión y eficiencia.

c) Redes NX

Un grafo es una estructura de datos formadas por nodos y aristas, la que es usada para representar relaciones entre objetos. Pueden emplearse para entrenar modelos de aprendizaje automático y realizar inferencias sobre los nodos, aristas o conjuntos de nodos (clústeres).

Los grafos se clasifican según la naturaleza de sus nodos y aristas. Por ejemplo, los grafos dirigidos tienen conexiones con una unión unidireccional, mientras que, en los grafos no dirigidos, los enlaces no tienen una sola dirección, sino más bien una conexión bidireccional (Carazo y Amat, 2023a).

a) Louvain

El algoritmo de Louvain detecta comunidades de nodos que están altamente interconectados dentro de la red. Su proceso incluye una fase de agrupamiento, donde los nodos se asignan a comunidades y posteriormente las comunidades se condensan en nuevos nodos para recalculan enlaces. Estas fases se repiten hasta que no hay mejoras significativas (Carazo y Amat, 2023b). Así, este método es útil para descubrir patrones estructurales en la red, facilitando el análisis de subgrupos relevantes y sus interacciones.

b) Modelos de entrenamiento probados y sus fundamentos teóricos

i. Regresión Lineal

La regresión lineal es un modelo que ajusta una relación lineal entre una variable dependiente y una o más variables independientes. Utiliza la ecuación.

$$y = w * x + b$$

Donde w representa la pendiente de la línea y b el término de desplazamiento vertical. Su propósito principal es ajustar la línea de mejor ajuste que minimice la suma de los errores cuadráticos entre los valores observados y los predichos (Bobadilla, J., 2020)

i. Árboles de decisiones y Gradient Boosting

Los árboles de decisión son modelos predictivos que recorren un “árbol” a través de caminos hasta alcanzar un nodo terminal y así poder generar la predicción, la cual está determinada mediante la media de las variables respuesta presentes en ese nodo. Estas características han convertido a este modelo en una herramienta muy utilizada en estadística y en aprendizaje automático (Amat 2020a).

El Gradient Boosting combina de manera secuencial modelos de árboles de decisión, donde cada árbol corrige los errores de su antecesor para minimizar la pérdida. La predicción final resulta de la combinación de todos los árboles, lo que lo convierte en un método eficiente para modelar relaciones complejas en diversos problemas (Amat 2020b). Con esto, el estudio se estructura en 5 etapas con pasos que se detallan a continuación:

Preprocesamiento de datos

El preprocesamiento de datos es una de las etapas más crítica, ya que permite que los datos estén en el formato adecuado, entregando calidad en a la información para su análisis. Utilizando herramientas como NumPy y Pandas, se realiza la manipulación de los datos, mientras que Seaborn y Matplotlib facilitan la visualización para identificar singularidades en los datos y así poder tomar las decisiones de limpieza más exhaustivas. Este paso garantiza que los datos sean consistentes y relevantes para el objetivo del estudio.

Los pasos efectuados fueron los siguientes:

- a) Combinar los archivos de texto que contienen los discursos con el registro correspondiente en la base de datos.
- b) Filtrar base de datos por región (América Latina y Caribe) y el periodo de tiempo (1998 al 2019) para analizar.
- c) Identificar el idioma de los discursos mediante modelos de detección.
- d) Limpieza de base de datos mediante eliminación de columnas innecesarias y filas duplicadas.

Análisis de texto

El análisis de texto permite la extracción de información de los discursos mediante técnicas de procesamiento de lenguaje natural (NLP). Desde la identificación de entidades relevantes hasta el análisis temático, utilizando herramientas como Name Entity Recognition (NER) y modelos de lematización.

- a) Creación de diccionarios y listas con palabras clave (países, gentilicios y diccionario de migración) para identificar en los discursos.
- b) Recopilación de entidades a partir análisis NER y de las palabras en los diccionarios para asegurar la mayor información.
- c) Procesamiento del corpus de discursos sin procesar (eliminación de caracteres especiales, tokenización, lematización).
- d) Extracción de frecuencias de términos.

Creación de redes

Las redes nos permiten comprender las relaciones entre las entidades (países) mencionadas en los discursos utilizando grafos para representar estas conexiones. Esto incluye métricas como centralidad, peso de las aristas y grado de conexiones, lo que facilita identificar actores clave y las relaciones más significativas en las narrativas. En el análisis realizado de discursos presidenciales, al ser el objetivo la evaluación de las conexiones entre países según menciones realizadas, es importante la dirección de la mención, permitiendo identificar los países que actúan como emisores principales de discursos sobre otros y cuáles son los receptores predominantes de estas menciones.

- a) Construcción de una red dirigida de menciones entre países basada en los resultados del NER, donde la dirección va desde el país del presidente que emite el discurso hacia el país mencionado.
- b) Ponderación de aristas por frecuencia de las menciones, donde el peso de la arista se obtiene mediante la normalización de la frecuencia de menciones de un país respecto al máximo valor en el data set, para indicar la intensidad relativa de la relación entre dos nodos.
- c) Eliminación de aristas y nodos con un bajo peso de arista que corresponden a países con conexiones débiles o poco significativa dentro del análisis discursivo.
- d) Cálculo de métricas para análisis de redes:
 - Centralidad: Identificar países (nodos) clave en las narrativas.
 - Centralidad de Grado: Mide cantidad de conexiones directas de un nodo.

- Centralidad de Intermediación: Mide cuantas veces el nodo actúa como intermediario en conexión con otros nodos
- Centralidad de Cercanía: Mide el promedio de distancia de un nodo a los demás nodos del grafo.
- In degree y out degree: Mide el grado de un nodo, contando las conexiones que tiene, considerando el peso, aristas de entradas (in-degree) y de salidas (out-degree).

Creación de comunidades

El análisis de comunidades busca segmentar los países de la red en grupos que compartan similitudes discursivas mediante indicadores como las métricas de redes, permitiendo identificar patrones discursivos entre países en la red global.

- a) Realizar un análisis de comunidades en la red de menciones mediante algoritmos como Louvain, para detectar grupos de nodos densamente conectados, que representen bloques discursivos.
- b) Incluir las métricas detectadas en la base de datos para una caracterización más precisa de la relevancia e influencia de los nodos en la red discursiva.

Predicción de nivel de populismo

La predicción de nivel de populismo permite identificar cómo factores afectan la aceptación de discursos. Utilizando modelos predictivos, se analizan variables para predecir los puntajes de populismo del discurso.

- a) Seleccionar las variables más relevantes mediante una matriz de correlación.

- b) Entrenar los modelos con un 80% de los datos disponibles y reservar el 20% restante para evaluación.
- c) Evaluar diferentes modelos de predicción utilizando métricas como error cuadrático medio (MSE) y coeficiente de determinación (R^2).

5. Resultados

A continuación, se presentan los resultados de cada una de las etapas del análisis, donde se puede observar que cada sección entrega información importante para la continuación de las siguientes fases.

5.1. Preprocesamiento de datos

Una vez insertados los discursos en la base de datos, se realizó el filtrado por región, seleccionando únicamente los discursos de América Latina en el periodo comprendido entre 1998 y 2019.

Se llevó a cabo la limpieza de la base de datos, eliminando columnas innecesarias como 'id', 'nota_populismo', 'region_wb', 'partido' y 'codificador', se identificó el idioma de los discursos utilizando modelos de detección de lenguaje y se filtraron aquellos escritos en español.

Se eliminaron filas duplicadas utilizando el nombre del archivo como referencia, ya que cada discurso contaba con múltiples registros asociados a las calificaciones individuales de distintos codificadores, manteniendo la columna promedio_discurso, que representa la nota promedio del nivel de populismo asignado al discurso.

El resultado de este preprocesamiento es la reducción de los datos a 273 discursos de 17 países (Fig. 3)

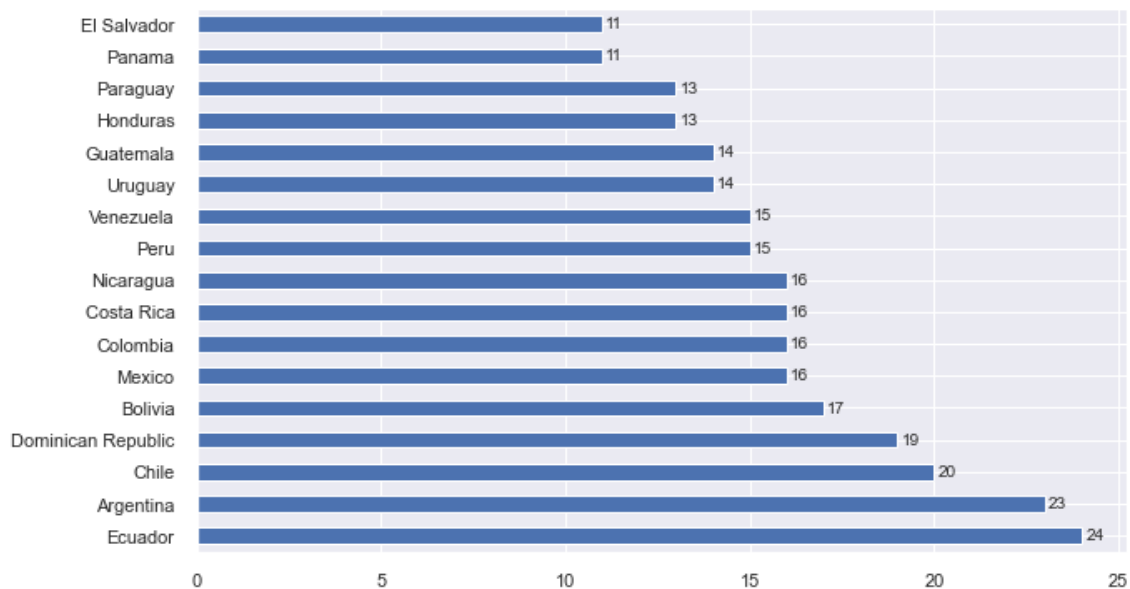


Fig. 3 Discursos por país

5.2. Resultados del análisis de texto

Utilizando la detección de entidades con NER y acompañándolo con los diccionarios creados para detectar países y palabras clave relacionadas con migración, se encontraron en total 1029 menciones de palabras del contexto sobre migración o extranjeros y un total de 887 menciones a otros países dentro de los discursos. Esta información es clave para la construcción de la red de menciones donde las conexiones representan las menciones y el peso que reflejan el nivel de populismo y la frecuencia de palabras clave sobre migración.

5.3. Construcción de red

La red se construyó utilizando menciones de países extraídas de los discursos mediante NER para armar los nodos, y los enlaces tienen peso según la variable 'frecuencia_paises', la cual fue normalizada (rango de 0 a 1) respecto al máximo valor en el data set.

De esta manera, para esta red se obtuvieron 39 nodos y 207 aristas, con una densidad inicial de 0.14, lo cual indica que la red no está interconectada en su totalidad, ya que existen una cantidad considerable de nodos que no se mencionan entre sí, lo cual se debe a la limitación del estudio, focalizado en la región de América Latina, ya que los nodos más lejanos corresponden a países de otras regiones.

La distribución de los pesos de las aristas (Tabla 3) revela que la mayoría de las conexiones se concentran en rangos bajos (0.0 - 0.5), mientras que los pesos superiores a 0.7 son significativamente menos frecuentes, exceptuando el rango de 1.0 a 2.0, que iguala a las menciones más bajas en frecuencia. De esta manera, se establece un corte de 0.3 en el peso de las aristas, reduciéndose a 97 conexiones, aumentando así ligeramente la modularidad de la red de 0.19 a 0.23, indicando una separación algo mejor de las comunidades, sin embargo, este cambio marginal sugiere que la estructura general de la red permanece similar. Este resultado refleja las limitaciones del diseño de la red, centrado en América Latina, donde las conexiones relevantes destacan en las narrativas de populismo y migración, facilitando un análisis más enfocado.

Tabla 3. Frecuencia de pesos en la red

Rango Peso	Frecuencia
0-0.1	59
0.1-0.2	25
0.2-0.3	26
0.3-0.4	10
0.4-0.5	14
0.5-0.6	12
0.6-0.7	14
0.7-0.8	7
0.8-0.9	3
0.9-1.0	2
1.0-2.0	32
2.0-3.0	2
3.0-4.0	1

5.4. Generación de comunidades

Utilizando el algoritmo de Louvain, tras eliminar aristas con pesos menores a 0.3, resulto en la identificación en la red de 3 comunidades (Fig. 4).

Las comunidades corresponden a (Tabla 4):

- a) Comunidad 0: Presenta la mayor conectividad (grado: 0.28) y un in-degree promedio elevado (3.84), lo que indica que son mencionados con frecuencia en los discursos. Aunque su intermediación es baja (0.01), actúan como una comunidad cohesionada. Incluye países latinoamericanos (Argentina, Chile, Bolivia, Paraguay y Uruguay) junto con naciones europeas y asiáticas (Norway, Spain, Austria e India).

- b) Comunidad 1: Tiene conectividad moderada (grado: 0.26) y destaca por su papel de puente en la red (intermediación: 0.02). Su out-degree promedio (3.29) sugiere que esta comunidad genera más menciones que las que recibe, reflejando un rol más activo en las narrativas. Agrupa países de América Latina, Norteamérica y Centroamérica (Colombia, Ecuador, Venezuela, México, Canadá, entre otros).
- c) Comunidad 2: Compuesta por países europeos, asiáticos y centroamericanos (Germany, Italy, Indonesia, El Salvador, entre otros), ocupa una posición periférica en la red. Su baja conectividad (grado: 0.119) y su limitada interacción (in-degree y out-degree cercanos a 1.68) indican que estos países son mencionados de forma más aislada en discursos específicos, con poca influencia en las narrativas globales.

En general, la modularidad de 0.23 indica una separación moderada entre comunidades, con las dinámicas más cohesionadas concentradas en la Comunidad 0, con países latinoamericanos como Argentina, Chile, y Bolivia. Mientras, la Comunidad 1 refleja una fuerte interacción entre países de América Latina y actores globales prominentes como Estados Unidos y Canada. La Comunidad 2, más periférica, incluye países europeos y centroamericanos como Alemania, Italia y El Salvador, con una menor conectividad y participación en las narrativas discursivas globales.

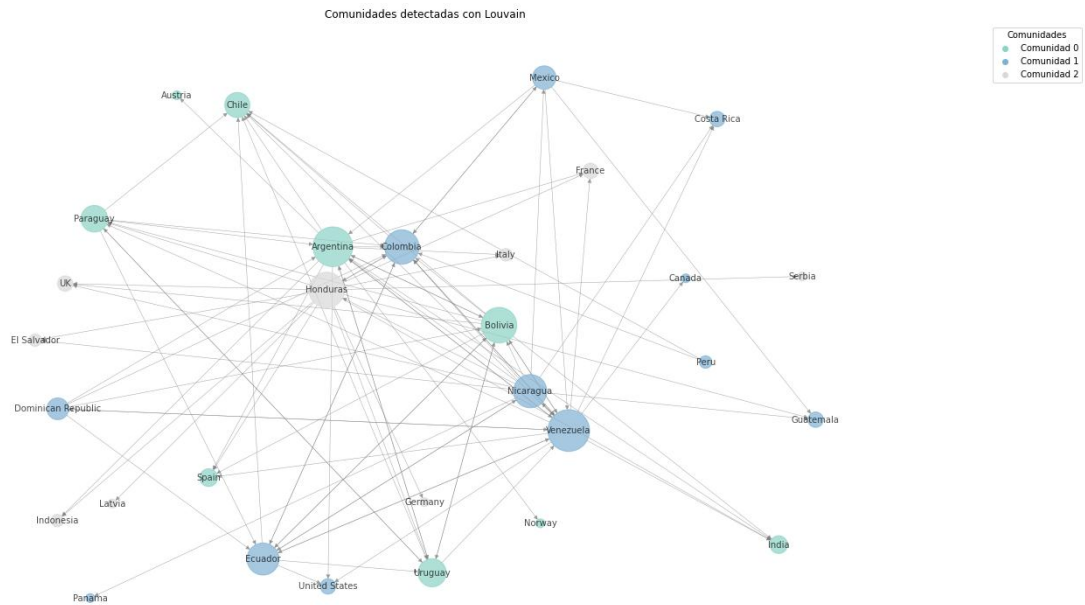


Fig. 4 Red de menciones de los países por comunidades detectadas

Tabla 4 Resultados comunidad mediante Louvain

Comunidad	Centralidad Grado	Centralidad intermediación	Centralidad Cercanía	Grado de entrada	Grado de salida
0	0.279693	0.011045	0.232838	3.835859	3.075758
1	0.258621	0.020363	0.201035	2.714015	3.289773
2	0.118774	0.007371	0.185302	1.679293	1.671717

5.5. Predicción de nivel de populismo

Finalmente, para generar un modelo que entregue una predicción del nivel de populismo, se seleccionaron variables que tuvieran una alta correlación (Fig. 5) evitando utilizar variables que de alguna manera están relacionadas directamente sobre esta, como ‘nota_promedio_periodo’, que tiene una alta correlación debido a que corresponde a el promedio de todos los discursos de un líder durante su mandato.

De esta manera, las variables seleccionadas son:

- ‘promedio_discurso_pais’: Representa la tendencia promedio de populismo en un país. Aunque podría considerarse parcialmente endógena, aporta un contexto esencial al reflejar características del país.
- ‘espectro_politico’: Indica la orientación ideológica del líder del discurso. Líderes en los extremos del espectro político tienden a ser más populistas, lo que hace de esta variable un predictor importante.
- ‘centralidad_grado’: Evalúa la prominencia de un país en la red a través de sus conexiones directas (in_degree y out_degree). Países más centrales pueden usar el populismo estratégicamente para consolidar liderazgo.
- ‘periodo’: Indica si el discurso pertenece al primer o segundo mandato presidencial. Esto permite capturar diferencias en la narrativa populista a lo largo del tiempo, como ajustes en el estilo de liderazgo o prioridades discursivas.
- ‘frecuencia_migracion’: Refleja la cantidad de menciones a temas migratorios en un discurso. Aunque su correlación es menor, mejora el modelo al añadir un componente temático crucial.

Estas variables fueron seleccionadas porque, juntas, ofrecen una combinación de contexto estructural (promedio_discurso_pais y espectro_politico), métricas de red (centralidad_grado) y factores temáticos y temporales (frecuencia_migracion y periodo).

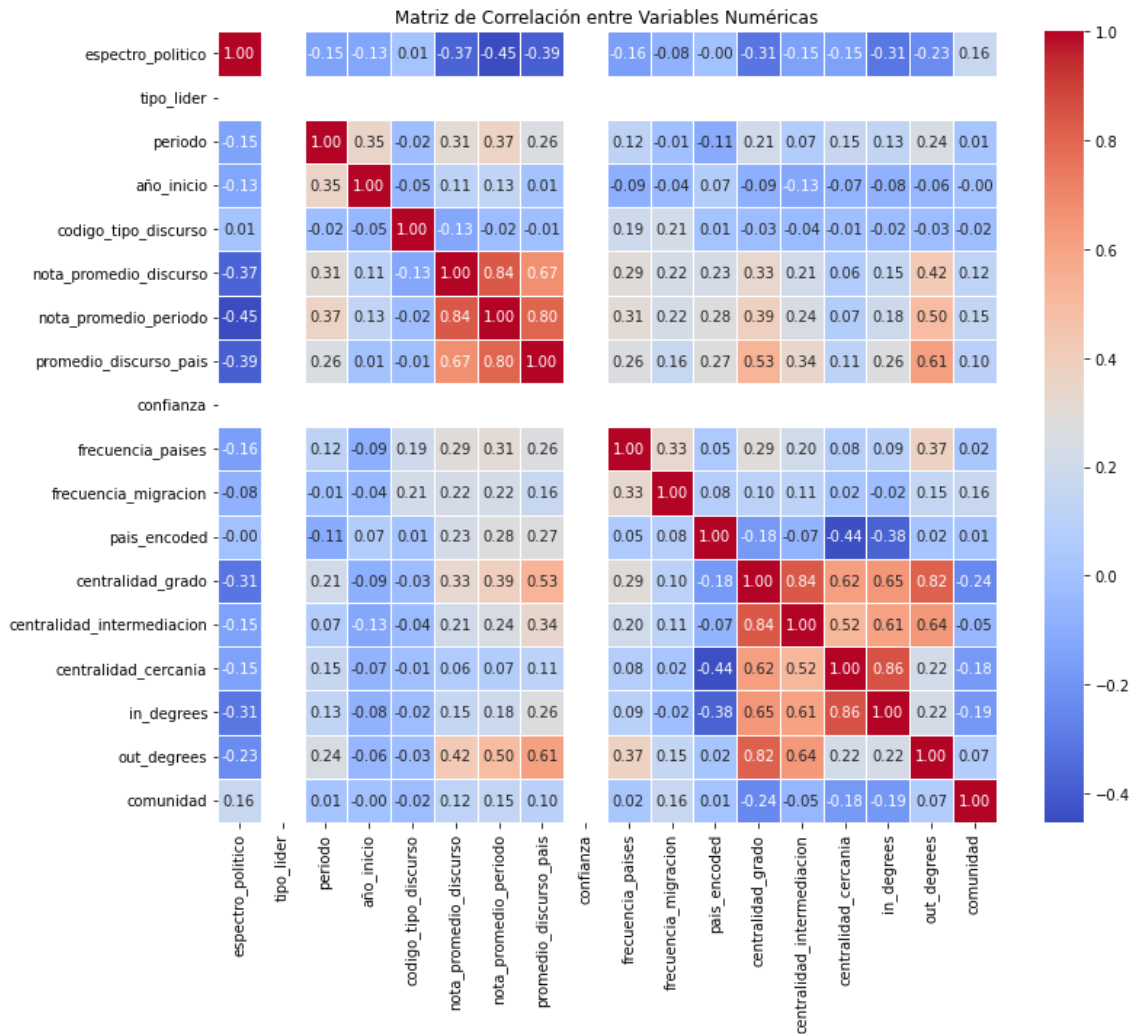


Fig. 5 Correlación entre variables

Se evaluaron diferentes modelos para la predicción, obteniendo como mejor resultado el modelo Gradient Boosting, con un MSE (Error cuadrático medio) de 0.1 y un R^2 de 0.63, lo que indica que el modelo logra un ajuste del 63% entre los valores reales y los predichos para el nivel de populismo (Fig. 7). El 37% de la variabilidad que no es explicada que puede deberse a la complejidad inherente al análisis del populismo, que puede depender de factores no considerados en el conjunto de variables utilizadas.

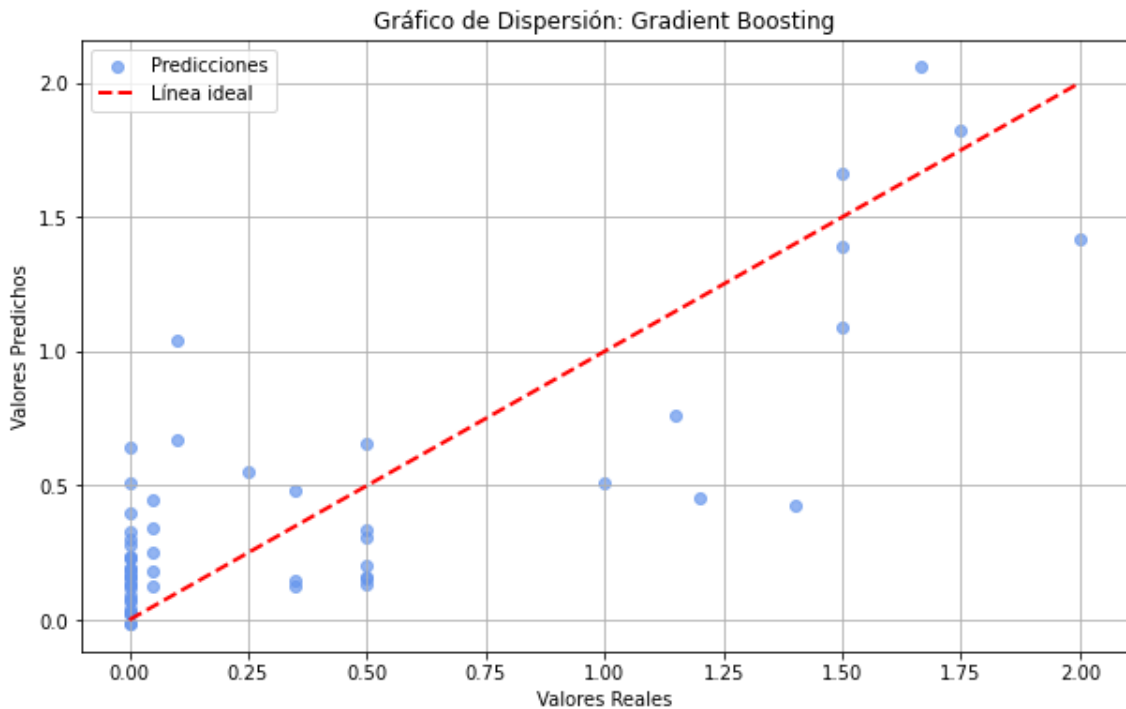


Fig. 6 Grafico de resultado predicción Regresión Lineal

5.6. Análisis Complementario

Como análisis complementario, los discursos se agruparon en terciles según la frecuencia de mención de temas migratorios. Luego, se calculó el nivel promedio de populismo en cada grupo, obteniendo valores de 0.36, 0.40 y 0.62, para los grupos de frecuencia baja (0-1), media (1-3) y alta (3-45), indicando que con mayor frecuencia de menciones migratorias tienden a exhibir niveles más altos de populismo (Fig. 8). Sin embargo, se debe destacar que los niveles de populismo en la base de datos están distribuidos de manera desigual, con un predominio de discursos con nivel de populismo 0 (121 discursos), seguidos por niveles moderados como 0.5 y 1.5, y los niveles más altos con menor frecuencia.

Esto sugiere que, aunque los niveles elevados de populismo están relacionados con temas migratorios, la mayoría de los discursos no son inherentemente populistas, pudiendo sugerir que las narrativas migratorias se utilizan estratégicamente en contextos específicos.

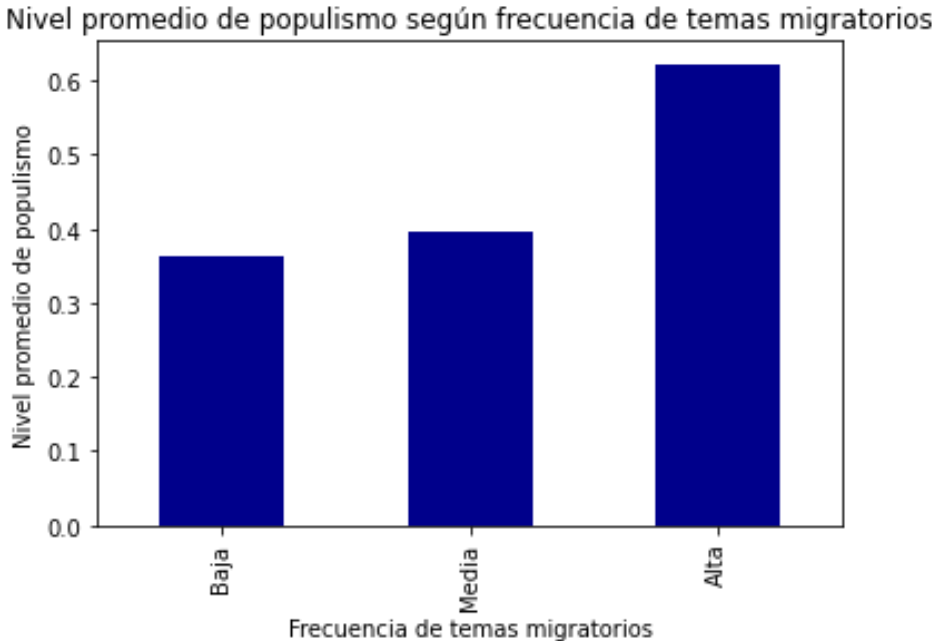


Fig. 7 Niveles promedio de populismo según la frecuencia de menciones migratorias

6. Conclusiones

Como se pudo observar en la investigación y análisis realizado, se abordaron dos aspectos principales: la generación de comunidades mediante el algoritmo Louvain y el desarrollo de un modelo predictivo para estimar el nivel de populismo de un discurso.

Con el uso de Named Entity Recognition (NER), el cual permitió identificar 887 menciones a países y 1029 menciones de palabras clave sobre migración, lo que proporcionó la base para construir una red de análisis dirigida. Esta red, evidenció que los discursos con altos niveles de populismo tienden a priorizar menciones a países vecinos o relevantes en temas migratorios. A través del algoritmo de Louvain, se identificaron 3 comunidades principales, y que tomando en consideración que los discursos analizados son exclusivamente de países de América Latina, lo que influye directamente en las comunidades detectadas, ya que la narrativa cohesionada de la Comunidad 0 refleja principalmente dinámicas regionales, mientras que las conexiones con actores globales en otras comunidades (como la Comunidad 2) se deben a menciones específicas en los discursos.

Además, se observó que discursos con mayor frecuencia de temas migratorios exhiben niveles más altos de populismo, con promedios de 0.36, 0.40 y 0.62 para frecuencias bajas, medias y altas, respectivamente. Sin embargo, los valores de populismo están distribuidos de manera desigual, predominando los discursos sin populismo (121 discursos). Esto sugiere que las narrativas migratorias pueden ser utilizadas estratégicamente en contextos específicos.

Así, el modelo predictivo desarrollado con Gradient Boosting alcanzó un desempeño sólido (R^2 de 0.63, MSE de 0.1), integrando variables clave como 'promedio_discurso_pais', 'espectro_politico', 'frecuencia_migracion' y 'centralidad_grado'. Aunque el 37% de la variabilidad no fue explicada, esto refleja la complejidad del populismo, influenciado por factores contextuales y retóricos posiblemente no considerados.

En conclusión, las narrativas migratorias pueden ser estratégicas en discursos populistas. Aunque el modelo capturo tendencias clave, la dispersión en los resultados sugiere la influencia de otros factores y la necesidad de ajustar parámetros o utilizar enfoques más robustos. Si bien, existe una relación entre populismo y menciones de temas sobre migración, la baja correlación indica que existen otros factores, debido a que los discursos populistas varían según la época, el país y el contexto global, adaptando sus temas a las inquietudes en cada momento, y la migración ha ganado mayor relevancia en los últimos años, resaltando la necesidad de más estudios.

6.1. Propuestas para trabajos futuros

- Incluir discursos en otros idiomas para obtener una base más robusta e identificar patrones globales y específicos por región.
- Usar técnicas de análisis de sentimiento para evaluar el tono emocional de los discursos relacionados con migración.
- Examinar cómo las narrativas en los discursos coinciden o divergen de los flujos migratorios reales.

Bibliografía

- [1] Amat, J. (2020a). Árboles de decisión con Python: regresión y clasificación. Ciencia de Datos. Recuperado de: https://cienciadedatos.net/documentos/py07_arboles_decision_python
- [2] Amat, J. (2020b). Gradient Boosting con Python. Ciencia de Datos. Recuperado de: https://cienciadedatos.net/documentos/py09_gradient_boosting_python
- [3] Bobadilla, J. (2020). Machine Learning y Deep Learning: Usando Python, Scikit y Keras. Bogotá.
- [4] Carazo, F. & Amat, J. (2023a). Introducción a grafos y redes con Python. Ciencia de Datos. Recuperado de: <https://cienciadedatos.net/documentos/pygml01-introduccion-grafos-redes-python>
- [5] Carazo, F. & Amat, J. (2023b). Análisis de redes con Python y NetworkX. Ciencia de Datos. Recuperado de: <https://cienciadedatos.net/documentos/pygml03-analisis-redes-python-networkx>
- [6] Correia, A. (2023). The rise of populist attitudes: using supervised machine learning to identify their main determinants.
- [7] Haidt, J., & Graham, J. (2007). When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not Recognize. En Social Justice Research (Vol. 20, Issue 1, pp. 98–116). Springer Science and Business Media LLC. <https://doi.org/10.1007/s11211-007-0034-z>

- [8] Hawkins, K. A., Aguilar, R., Castanho Silva, B., Jenne, E. K., Kocijan, B., & Rovira Kaltwasser, C., (2019). Measuring Populist Discourse: The Global Populism Database. Disponible en: https://populism.byu.edu/App_Data/Publications/Global%20Populism%20Database%20Paper.pdf
- [9] Martínez, A. (2023). Populismo en América Latina: el eterno retorno. *Ciencia Política*, 19–44. MLA
- [10] McAuliffe, M. & Oucho, L. (2024). World Migration Report 2024. International Organization for Migration (IOM).
- [11] Stuardo, M., Soler, S., & Riera, M. (2021). Discursos políticos y mediáticos contemporáneos sobre los inmigrantes. *Publicación del Instituto Universitario de Estudios Sobre Migraciones*, (52), 31-57. Disponible en: <https://revistas.comillas.edu/index.php/revistamigraciones/article/view/12935/14982>
- [12] Ulinskaitė, J., & Pukelis, L. (2021). Identifying Populist Paragraphs in Text: A machine-learning approach. Cornell University, USA. <https://doi.org/10.48550/arxiv.2106.03161>