



Derecha fragmentada y un enemigo compartido

Análisis textual longitudinal de la constación discursiva en las elecciones presidencial de 2025 en Chile

Tesis para optar al grado de
Magíster en Ciencia de Datos

Matías Deneken

Centro de Datos e Inteligencia Artificial - Universidad de Concepción

Profesores Guias:

Dr. Carlos Navarrete
Dr. Marcela Parada

Universidad de Concepción

Enero 2026

Agradecimientos

Esta tesina no habría sido posible por la beca de excención de arancel por el Premio Universidad de Concepción, según consta el decreto DIRPOST-021-24.

Agradezco a mis compañeras y compañeros de generación con los cuáles compartí durante este proceso. En particular a Javiera Baeza y Florencia Pampaloni, con quién constituyó un excelente equipo de trabajo desde la distancia. Sin este trabajo colaborativo la experiencia en el MACI habría sido significativamente más compleja.

Presentación

Contenido de presentación/introducción general

Table of contents

Agradecimientos	ii
Presentación	iii
1 Introducción	1
1.1 Presentación del problema de investigación	1
1.1.1 Justificación y relevancia del estudio	2
1.2 1.3. Preguntas de investigación	3
1.3 1.4. Objetivos de la investigación	3
1.3.1 Objetivo general	3
1.3.2 Objetivos específicos	4
2 Presentación del Problema	5
2.1 Problema	5
2.2 Acercamiento, actores y dimensiones	5
2.3 Descripción de la propuesta	5
2.4 MVP y objetivos	5
2.5 Validación	5
3 Diseño metodológico	6
3.1 Recolección de datos: web scraping mediante API	7
3.1.1 Procedimiento técnico	7
3.1.2 Período temporal	7
3.1.3 Estrategia de scraping incremental	8
3.1.4 Filtrado temático	8

3.1.5	Figuras políticas de interés	9
3.2	Consideraciones éticas	10
3.3	Técnicas de análisis	10
3.3.1	Análisis de sentimiento y emociones	10
3.3.2	Modelado de tópicos	11
3.3.3	Análisis de redes de interacción	11
3.3.4	Análisis temporal	11
3.4	Volumen y estructura del corpus final	11
4	Depuración de Datos	13
5	Exploración de Datos	14
6	Modelado de Datos	15
6.1	Clustering K-Means	15
6.2	Clustering DBSCAN	15
6.3	Clustering Jerárquico	15
6.4	Determinación del modelo más óptimo	15
7	Interpretación de Resultados	16
7.1	Interpretación de componentes principales	16
7.2	Interpretación de clusters	16
7.3	Validación de los resultados	16
8	Conclusión	17
9	Referencias	18
10	Referencias	19

List of Figures

List of Tables

1 Introducción

1.1 Presentación del problema de investigación

Las elecciones presidenciales chilenas de 2025 se desarrollan en un contexto de profunda reconfiguración del sistema de partidos y de transformación de las condiciones bajo las cuales se produce la comunicación política. Tras el ciclo de movilizaciones iniciado en octubre de 2019, los procesos constituyentes fallidos de 2022 y 2023, y la creciente fragmentación de las coaliciones tradicionales, el escenario electoral chileno exhibe niveles de incertidumbre, volatilidad y polarización inéditos en la historia reciente del país.

En este contexto, las plataformas digitales han adquirido un protagonismo sin precedentes como espacios de debate político, formación de opinión y articulación de identidades colectivas. Lo que ocurre en redes sociales ya no constituye un reflejo secundario de dinámicas políticas que se dirimen en otros ámbitos, sino un terreno constitutivo donde se disputan los términos del conflicto, se movilizan emociones políticas y se construyen las fronteras simbólicas que delimitan aliados y adversarios.

Una de las transformaciones más significativas del período reciente es la reconfiguración del campo de las derechas chilenas. Donde antes existía una coalición relativamente cohesionada —heredera de la transición democrática y articulada en torno a partidos como la UDI y Renovación Nacional—, hoy convergen y compiten al menos tres corrientes distinguibles: una derecha tradicional que busca renovar su apelación desde la moderación y la competencia técnica; una derecha radical-conservadora que articula posiciones morales restrictivas con retórica antielitista; y una derecha libertaria de perfil disruptivo que desafía las convenciones comunicativas establecidas. La competencia entre

estas corrientes, y su eventual convergencia frente a adversarios comunes, constituye uno de los fenómenos más relevantes —y menos estudiados— de la política chilena contemporánea.

1.1.1 Justificación y relevancia del estudio

Esta investigación se justifica en función de tres órdenes de consideraciones: empíricas, teóricas y metodológicas.

Desde el punto de vista **empírico**, el ciclo electoral chileno de 2025 ofrece una oportunidad única para observar en tiempo real las dinámicas de competencia intra-derecha en un sistema político en transformación. La coexistencia de tres candidaturas de derecha con perfiles ideológicos y estrategias comunicativas diferenciadas —Evelyn Matthei, José Antonio Kast y Johannes Kaiser— permite analizar comparativamente cómo distintas corrientes construyen sus identidades, movilizan a sus bases y se posicionan frente a adversarios internos y externos. La presencia de una candidatura de izquierda asociada al Partido Comunista —Jeannette Jara— añade una dimensión adicional al análisis, habilitando el estudio de los procesos de convergencia discursiva que emergen cuando actores previamente enfrentados deben articular un frente común.

Desde el punto de vista **teórico**, la investigación contribuye a debates contemporáneos sobre comunicación política digital, polarización afectiva y construcción de antagonismos. Si bien existe abundante literatura sobre el rol de las redes sociales en procesos electorales, la mayor parte de esta investigación se ha concentrado en contextos del Norte Global, particularmente Estados Unidos y Europa Occidental. Los estudios sobre América Latina, aunque crecientes, han tendido a focalizarse en casos de alta visibilidad internacional —Brasil con Bolsonaro, Argentina con Milei— sin atender suficientemente a las variaciones nacionales y a las dinámicas de competencia al interior de familias ideológicas. Esta tesis busca contribuir a llenar ese vacío, ofreciendo un análisis detallado de las especificidades del caso chileno.

Desde el punto de vista **metodológico**, la investigación propone una aproximación que integra técnicas de ciencia social computacional con análisis interpretativo del discurso político. Esta combinación permite procesar volúmenes de datos que excederían las capacidades del análisis cualitativo tradicional, sin sacrificar la atención a las dimensiones semánticas y pragmáticas que resultan cruciales para comprender la comunicación política. La elección de Reddit como campo de observación

—una plataforma subexplorada en el contexto latinoamericano— ofrece además la posibilidad de acceder a conversaciones políticas con características estructurales distintivas que enriquecen el análisis.

1.2 1.3. Preguntas de investigación

La pregunta central que orienta esta investigación es:

¿Cómo se configura y evoluciona la conversación política en Reddit durante la campaña presidencial chilena de 2025?

Esta pregunta general se desagrega en las siguientes interrogantes específicas:

1. ¿Qué patrones de encuadre (*framing*), movilización emocional y construcción de adversarios estructuran el debate político en los subreddits chilenos durante el período electoral?
2. ¿Cómo varían temporalmente las estrategias discursivas de ataque, defensa y demarcación de fronteras identitarias a lo largo de las distintas fases de la campaña?
3. ¿Qué dinámicas de competencia, diferenciación y antagonismo se observan entre las tres corrientes de derecha (tradicional, radical-conservadora y libertaria) durante la primera vuelta?
4. ¿Cómo se produce la convergencia discursiva en torno al Partido Comunista como adversario común durante la segunda vuelta, y qué tensiones persisten bajo esta convergencia aparente?

1.3 1.4. Objetivos de la investigación

1.3.1 Objetivo general

Analizar la configuración y evolución de la conversación política en Reddit durante la campaña presidencial chilena de 2025, identificando los patrones de encuadre, emoción y construcción de adversarios que estructuran el debate entre las distintas corrientes ideológicas.

1.3.2 Objetivos específicos

1. Caracterizar los marcos interpretativos (*frames*) predominantes en el discurso de cada corriente ideológica y su variación a lo largo del ciclo electoral.
2. Identificar los patrones de movilización emocional asociados a cada candidatura, distinguiendo emociones de valencia positiva, negativa y las combinaciones que configuran repertorios afectivos específicos
3. Analizar las estrategias de construcción de adversarios, atendiendo tanto a la definición de enemigos externos (inter-bloque) como a las dinámicas de diferenciación interna (intra-bloque).
4. Examinar la evolución temporal del debate, identificando puntos de inflexión, eventos catalizadores y patrones de activación discursiva asociados a momentos específicos de la campaña.
5. Comparar las dinámicas discursivas entre los dos subreddits analizados (r/chile y r/RepublicadeChile), identificando diferencias en composición ideológica, estilos argumentativos y patrones de interacción.

2 Presentación del Problema

[Contenido de la presentación del problema]

2.1 Problema

[Descripción del problema]

2.2 Acercamiento, actores y dimensiones

[Descripción del acercamiento, actores y dimensiones]

2.3 Descripción de la propuesta

Descripción de la propuesta

2.4 MVP y objetivos

[Descripción del MVP y objetivos]

2.5 Validación

[Descripción de la validación]

3 Diseño metodológico

Esta investigación adopta un diseño **longitudinal observacional** basado en el análisis de datos digitales generados por usuarios (*user-generated content*). El carácter longitudinal permite capturar la evolución temporal del debate político a lo largo del ciclo electoral, identificando patrones de variación, puntos de inflexión y dinámicas de cambio. El carácter observacional implica que los datos son recolectados sin intervención experimental, preservando las características naturales de la interacción en las plataformas estudiadas.

Metodológicamente, la investigación se inscribe en el campo emergente de la **ciencia social computacional** (*computational social science*), que integra técnicas de procesamiento de datos a gran escala con los marcos teóricos y las preguntas sustantivas de las ciencias sociales (Lazer et al., 2009; Salganik, 2019). Esta aproximación permite analizar volúmenes de datos textuales que excederían las capacidades del análisis cualitativo tradicional, mientras mantiene un compromiso con la interpretación teóricamente informada de los patrones identificados.

El diseño combina componentes **descriptivos**, orientados a caracterizar la conversación política en Reddit durante el período estudiado, con componentes **explicativos**, orientados a identificar los factores y mecanismos que dan cuenta de los patrones observados. Esta combinación responde a la naturaleza de las preguntas de investigación, que demandan tanto una cartografía del terreno empírico como una comprensión de las dinámicas que lo estructuran.

3.1 Recolección de datos: web scraping mediante API

3.1.1 Procedimiento técnico

Los datos fueron extraídos mediante **web scraping** utilizando la API pública de Reddit a través de la librería PRAW (*Python Reddit API Wrapper*). Esta aproximación permite acceder de forma estructurada y sistemática al contenido público de la plataforma, respetando los límites de uso establecidos por Reddit y garantizando la trazabilidad y reproducibilidad del proceso de recolección.

El scraping se realizó sobre los dos subreddits principales de la comunidad chilena:

- **r/chile**: Comunidad generalista sobre temas chilenos, con mayor volumen de usuarios y diversidad temática.
- **r/RepublicadeChile**: Comunidad alternativa surgida como escisión de r/chile, caracterizada por una composición ideológica distintiva y políticas de moderación más permisivas.

Estos espacios fueron seleccionados por constituir los principales foros de discusión política chilena en la plataforma, caracterizados por participación activa, diversidad de perspectivas ideológicas y debates sostenidos sobre temas de contingencia nacional.

3.1.2 Período temporal

La recolección de datos abarcó **cinco meses**, desde el **1 de agosto hasta el 31 de diciembre de 2024**. Este período temporal captura diferentes fases del ciclo electoral presidencial chileno de 2025:

- **Agosto-octubre 2024**: Fase de emergencia de candidaturas, posicionamiento inicial y primeras definiciones programáticas.
- **Noviembre-diciembre 2024**: Período de escalamiento de campaña, consolidación electoral y estructuración del debate público.

La ventana temporal de cinco meses permite observar la evolución dinámica del debate político, identificar patrones temporales de activación discursiva y capturar tanto tendencias sostenidas como

picos de discusión asociados a eventos políticos específicos del período (debates televisados, declaraciones públicas, controversias mediáticas).

3.1.3 Estrategia de scraping incremental

El proceso de recolección fue **incremental y acumulativo**, ejecutándose en múltiples iteraciones programadas que:

1. **Extracción:** Recuperaban los posts más recientes disponibles en cada subreddit (límite de 250-1000 posts por ejecución según configuración).
2. **Expansión:** Descargaban el árbol completo de comentarios de cada post relevante mediante `replace_more(limit=None)` para capturar todos los niveles de anidación conversacional.
3. **Acumulación:** Integraban nuevos datos con extracciones previas en archivos CSV, Parquet y Excel.
4. **Deduplicación:** Eliminaban registros duplicados mediante identificadores únicos generados con hashes SHA-256 basados en: `post_id + timestamp + autor + contenido textual`.

Este diseño iterativo permitió maximizar la cobertura temporal sin exceder los límites de la API, capturar discusiones en desarrollo a lo largo de los cinco meses y construir un corpus longitudinal que refleja la evolución dinámica del debate político.

3.1.4 Filtrado temático

Para cada publicación dentro del rango temporal, se trajeron dos niveles de contenido textual:

Nivel de publicación (posts):

- Título (*title*): Encabezado del post que condensa el tema o posición principal.
- Cuerpo (*selftext*): Texto completo de la publicación cuando el formato lo incluye.
- Metadatos: autor, fecha de creación, puntuación, número de comentarios, flair (etiqueta temática).

Nivel de comentarios:

- Cuerpo del comentario (*comment_body*): Texto completo de cada comentario en el árbol de discusión.
- Metadatos: autor, fecha de creación, puntuación, relación con post padre e identificador del comentario al que responde.

Ambos niveles de texto fueron procesados mediante **expresiones regulares** que identifican menciones de figuras políticas específicas, manejando variaciones ortográficas comunes (presencia/ausencia de acentos, nombres completos vs. apellidos, apodos políticos, errores tipográficos frecuentes). Este filtrado temático permite construir un corpus focalizado en la discusión sobre las candidaturas presidenciales de interés.

3.1.5 Figuras políticas de interés

El filtrado temático se centró en cuatro candidaturas representativas de las distintas corrientes ideológicas:

Candidatura	Corriente ideológica	Patrones de búsqueda
Evelyn Matthei	Derecha tradicional	“matthei”, “matei”, “evelyn”
José Antonio Kast	Derecha radical-conservadora	“kast”, “jose antonio kast”, “jak”
Johannes Kaiser	Derecha libertaria	“kaiser”, “johannes”, “jkaiser”
Jeannette Jara	Izquierda (PC)	“jara”, “jeannette”, “jeanette”

Esta selección permite analizar tanto la **competencia intra-derecha** entre tres corrientes (tradicional, republicana, libertaria) como la **convergencia discursiva** frente a un adversario ideológico común en segunda vuelta (el Partido Comunista).

3.2 Consideraciones éticas

La investigación se rige por principios éticos establecidos para la investigación con datos digitales (Salganik, 2019; Zimmer & Kinder-Kurlanda, 2017):

Consentimiento y datos públicos: Los datos recolectados provienen de publicaciones y comentarios realizados en espacios públicos de la plataforma, donde los usuarios tienen expectativas razonables de que su contenido será visible para audiencias amplias. No se accedió a contenido privado, mensajes directos ni información restringida.

Anonimización: Aunque los nombres de usuario en Reddit son seudónimos, el análisis no reporta identificadores individuales ni busca vincular cuentas con identidades reales. Los ejemplos citados en el análisis son anonimizados cuando resulta necesario.

Minimización de daños: La investigación no interviene en las comunidades estudiadas ni expone a sus miembros a riesgos adicionales. El objetivo es comprender dinámicas colectivas, no escrutinizar comportamientos individuales.

Reproducibilidad: Los scripts de recolección y procesamiento se documentan para permitir la replicación del estudio, respetando los términos de uso de la API de Reddit.

3.3 Técnicas de análisis

3.3.1 Análisis de sentimiento y emociones

Para identificar los patrones emocionales del corpus, se implementó análisis de sentimiento mediante modelos de procesamiento de lenguaje natural entrenados para español. Se empleó una aproximación multidimensional que distingue entre:

- **Polaridad:** Valencia positiva, negativa o neutra del contenido.
- **Intensidad:** Grado de activación emocional.
- **Emociones discretas:** Identificación de emociones específicas (alegría, tristeza, miedo, ira, sorpresa, asco) cuando los modelos lo permiten.

3.3.2 Modelado de tópicos

Para identificar los marcos interpretativos y temas estructurantes del debate, se implementaron técnicas de **modelado de tópicos** (*topic modeling*). Estas técnicas permiten descubrir patrones temáticos latentes en grandes corpus textuales, identificando agrupaciones de palabras que tienden a co-ocurrir y que pueden interpretarse como “temas” o, en el marco teórico de esta investigación, como indicadores de *frames*.

3.3.3 Análisis de redes de interacción

Para examinar las dinámicas interaccionales, se construyeron **redes de respuesta** donde los nodos representan usuarios y las aristas representan relaciones de respuesta (comentario a comentario). Este análisis permite identificar patrones de interacción entre usuarios de distintas orientaciones ideológicas, la existencia de subcomunidades discursivas y las dinámicas de confrontación o cooperación.

3.3.4 Análisis temporal

El diseño longitudinal permite implementar análisis de **series temporales** que examinan cómo varían los indicadores de interés a lo largo del período estudiado. Esto incluye la identificación de tendencias, ciclos, puntos de quiebre y la correlación entre eventos externos (debates, declaraciones, escándalos) y cambios en los patrones discursivos.

3.4 Volumen y estructura del corpus final

Al cierre de la recolección (31 de diciembre de 2024), el corpus construido contiene:

- Todos los posts publicados en r/chile y r/RepublicadeChile durante el período que mencionan al menos una figura política de interés.
- Todos los comentarios dentro de esos posts que mencionan a dichas figuras.
- Estructura temporal completa desde el 1 de agosto al 31 de diciembre de 2024.

- Metadatos asociados que permiten análisis longitudinal, de red y de interacción.

Este diseño metodológico permite observar la configuración de la conversación política en Reddit y su evolución durante la campaña presidencial chilena de 2025, con foco específico en patrones de enmarcamiento (*framing*), movilización emocional y construcción de adversarios a lo largo del ciclo electoral.

4 Depuración de Datos

5 Exploración de Datos

6 Modelado de Datos

[Contenido de modelado de datos]

6.1 Clustering K-Means

[Contenido sobre clustering K-Means]

6.2 Clustering DBSCAN

[Contenido sobre clustering DBSCAN]

6.3 Clustering Jerárquico

[Contenido sobre clustering jerárquico]

6.4 Determinación del modelo más óptimo

[Contenido sobre determinación del modelo más óptimo]

7 Interpretación de Resultados

[Contenido de interpretación de resultados]

7.1 Interpretación de componentes principales

[Contenido sobre interpretación de componentes principales]

7.2 Interpretación de clusters

[Contenido sobre interpretación de clusters]

7.3 Validación de los resultados

[Contenido sobre validación de resultados]

8 Conclusión

[Contenido de la conclusión]

9 Referencias

10 Referencias

[Las referencias se generan automáticamente desde el archivo bibliography.bib]