

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



“ANÁLISIS DE LOS HECHOS Y TENDENCIAS QUE
DETERMINARON EL RESULTADO DEL PLEBISCITO
CONSTITUCIONAL DE 2022, MEDIANTE PERIODISMO
DE DATOS”

SALVADOR DARÍO FUENTES AZÓCAR

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: José Luis Martí Lara
Profesor Correferente: Claudio Tapia Figueroa

Octubre - 2023

DEDICATORIA

Le dedico esta memoria a mi tío Yoyo, que en paz descansa, por haber sido la persona que me inspiró a conocer más sobre el mundo y el querer comprender el funcionamiento de las cosas, en especial la tecnología, finalmente llevándome a estudiar la hermosa carrera de Ingeniería Civil en Informática.

AGRADECIMIENTOS

Quiero, en primer lugar, agradecerle a mi familia, por el apoyo y confianza incondicionales que he recibido de ellos a lo largo de toda mi vida. Sin su compañía difícilmente habría llegado hasta este punto.

En segundo lugar, quiero agradecerle a mis compañeros de trabajo en Aramark, por mostrar su interés en esta memoria y motivarme a continuar con dicho estudio.

A mi profesor guía, José Luis Martí, por permitirme realizar una memoria que combinó mis dos grandes aficiones y por ayudarme en cada etapa del proceso.

A mi amiga Maga, cuya reaparición en mi vida me ayudó a superar uno de los momentos más difíciles que he vivido, quien también se ofreció a ayudarme a redactar esta memoria.

Finalmente, pero no menos importante, quiero agradecerle a mis amigos de esta universidad, quienes merecen un espacio personal para cada uno de ellos, por ser quienes me ayudaron y empujaron en cada dificultad que pasé a lo largo de mi carrera.

Quiero agradecerles a Sebastián Calderón y a Jaime Schweir, mis primeros amigos en esta universidad, con quienes pasé y sigo pasando grandes momentos en la USM.

A Manuel Valenzuela, Daniel Magaña, Sebastián Sepúlveda y Emiliano Santana por ser el grupo de amigos más longevo y cercano que he tenido en mi vida universitaria, cada uno de ustedes son personas maravillosas con futuros brillantes que sin duda han marcado mi vida y esta memoria.

A mis amigas Nazareth Díaz y Jael Toledo, por una amistad que se ha ido fortaleciendo con los años y que me han acompañado incansablemente durante las etapas de este camino.

A Nicole Linares, Fernanda Cerda, Isidora Ubilla, Nicolás Pujante, Cristian Calderón, Daniel López, Alfredo Llanos, Luis Corrales y Astrid Chirinos, a quienes conocí en diferentes etapas de mi carrera universitaria, y que siempre es un agrado compartir con ellos.

RESUMEN

Resumen— A partir del año 2020, Chile vivió una gran cantidad de procesos electorarios, siendo los más importantes el plebiscito constitucional de 2020, las elecciones presidenciales del 2021 y, finalmente, el plebiscito constitucional de 2022. Con la victoria contundente del Apruebo en 2020 y el triunfo de Gabriel Boric en 2021, era razonable asumir que una tendencia parecida iba a darse en el 2022. Sin embargo, el Rechazo logró doblegar al Apruebo con un 62 % de las votos, una diferencia que nadie logró prever. Este trabajo buscará encontrar aquellos hechos y tendencias que llevaron a este sorpresivo resultado mediante el periodismo de datos, una disciplina que incorpora conceptos del periodismo y la informática para generar un panorama más amplio en el estudio sobre temas sociales. Así, será necesaria la creación de una fuente de datos utilizando *web scraping* en Twitter para su posterior análisis, realizando procesamiento de texto y análisis de sentimiento a través de *machine learning*, con el fin de poder identificar los tópicos más recurrentes en el debate constitucional, la percepción de los votantes respecto a éstos y los personajes que lograron influenciar a una mayor cantidad de votantes con sus declaraciones.

Palabras Clave— Plebiscito constitucional, periodismo de datos, *web scraping*, *machine learning*, análisis de sentimiento.

ABSTRACT

Abstract— From the year 2020 onwards, Chile experienced a significant number of electoral processes, with the most notable being the constitutional plebiscite of 2020, the presidential elections of 2021, and finally, the constitutional plebiscite of 2022. With the resounding win of “Apruebo” in 2020 and Gabriel Boric’s triumph in 2021, it was reasonable to assume a similar trend in 2022. However, “Rechazo” managed to overcome “Apruebo” with 62 % of the votes, a difference that no one foresaw. This work aims to identify the events and trends that led to this surprising outcome through data journalism, a discipline that incorporates concepts from journalism and computer science to provide a broader perspective in the study of social issues. Thus, the creation of a data source using web scraping on Twitter will be necessary for subsequent analysis. This analysis will involve text processing and sentiment analysis through machine learning to identify the most recurring topics in the constitutional debate, the voters’ perception of these topics, and the personalities who successfully influenced a greater number of voters with their statements.

Keywords— Constitutional plebiscite, data journalism, *web scraping*, *machine learning*, sentiment analysis.

ÍNDICE DE CONTENIDOS

RESUMEN	IV
ABSTRACT	IV
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	IX
INTRODUCCIÓN	1
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	2
1.1 Escenario electoral de Chile entre el 2020 y el 2022	2
1.2 Objetivo de la solución	4
1.2.1 Objetivo general	4
1.2.2 Objetivos específicos	4
CAPÍTULO 2: MARCO CONCEPTUAL	5
2.1 Periodismo de datos	5
2.2 Metodologías para el análisis de datos	6
2.3 <i>Web scraping</i>	7
2.4 Análisis de sentimiento	8
2.5 Métodos de validación para algoritmos de análisis de sentimiento basados en <i>machine learning</i>	14
2.6 Visualización de datos	16
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	17
3.1 Elección de la metodología	17
3.2 Selección de los datos	18
3.3 Preprocesamiento de datos	22
3.4 Transformación de datos	23
3.5 Análisis de Sentimiento, Emoción y Detección de Discurso de Odio (PySentimiento)	24
3.5.1 Análisis de sentimiento	26
3.5.2 Análisis de emociones	47
3.5.3 Detección de discurso de odio	50
3.6 Interpretación de resultados	56
3.6.1 Distribución del sentimiento	57
3.6.2 Evolución del sentimiento a través del tiempo	59
3.6.3 Comparación de la evolución del sentimiento a través del tiempo	61
3.6.4 Positividad, neutralidad y negatividad	62
3.6.5 <i>Hashtags</i> y análisis de sentimiento	64
3.6.6 Los actores del debate y su rol e influencia en este	68

3.6.7 Distribución y evolución de las emociones	74
3.6.8 <i>Hashtags</i> y análisis de emociones	76
3.6.9 Evolución de la presencia de discurso de odio y el impacto de la plurinacionalidad e interculturalidad	79
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN	82
4.1 Metodología de validación	82
4.1.1 Grupo de validación	82
4.1.2 Cuestionario de entrada	82
4.1.3 Cuestionario de evaluación	83
4.1.4 Cuestionario de salida	84
4.2 Resultados de la validación	84
4.2.1 Cuestionario de entrada	84
4.2.2 Cuestionario de evaluación	85
4.2.3 Cuestionario de salida	90
CONCLUSIONES	91
ANEXOS	96
REFERENCIAS BIBLIOGRÁFICAS	97

ÍNDICE DE FIGURAS

1	Evolución de las encuestas del plebiscito constitucional de Chile de 2022. . . .	3
2	Anatomía de un <i>tweet</i>	8
3	Representación de un hiper plano en una SVM.	11
4	Estructura del análisis de sentimiento para <i>tweets</i> usando <i>machine learning</i> . . .	12
5	Ejemplo de una matriz de confusión.	14
6	Ejemplo de una Curva ROC.	15
7	Matriz de confusión para análisis de sentimiento.	28
8	Distribución del sentimiento en el <i>dataset</i> sin respuestas.	57
9	Distribución del sentimiento en el <i>dataset</i> con respuestas.	58
10	Evolución del sentimiento a través del tiempo en el <i>dataset</i> sin respuestas. . .	59
11	Evolución del sentimiento a través del tiempo en el <i>dataset</i> con respuestas. . .	60
12	Comparación de la evolución del sentimiento a través del tiempo.	61
13	Distribución de la positividad, neutralidad y negatividad.	62
14	Evolución de la positividad, neutralidad y negatividad a través del tiempo. . . .	63
15	<i>Hashtags</i> más recurrentes durante los meses de enero y septiembre de 2022. . .	64
16	<i>Hashtags</i> más recurrentes durante los meses de enero y septiembre de 2022, clasificados por sentimiento.	66
17	<i>Hashtags</i> más recurrentes y sus frecuencias durante los meses de enero y septiembre de 2022, clasificados por sentimiento.	67
18	Top 10 constituyentes más mencionados en Twitter, clasificados por el sentimiento vertido en sus menciones.	69
19	Top 10 personajes públicos no constituyentes más mencionados en Twitter, clasificados por el sentimiento vertido en sus menciones.	70
20	Top 10 constituyentes más mencionados en Twitter, posicionados de acuerdo al sentimiento vertido en sus menciones y su posición política.	71

21	Top 10 personajes públicos no constituyentes más mencionados en Twitter, posicionados de acuerdo al sentimiento vertido en sus menciones y su posición política.	73
22	Distribución de las emociones.	74
23	Evolución de las emociones a través del tiempo.	75
24	<i>Hashtags</i> más recurrentes durante los meses de enero y septiembre de 2022, clasificados por emoción.	77
25	<i>Hashtags</i> más recurrentes y sus frecuencias durante los meses de enero y septiembre de 2022, clasificados por emoción.	78
26	Evolución de la presencia de discurso de odio a través del tiempo.	79
27	Evolución de la presencia de discurso de odio en <i>tweets</i> relacionados a la plurinacionalidad e interculturalidad a través del tiempo.	80
28	Comparación de la presencia de discurso de odio a través del tiempo entre la presencia y ausencia de la plurinacionalidad e interculturalidad.	81
29	Evolución de la aprobación de Gabriel Boric durante el año 2022.	96

ÍNDICE DE TABLAS

1	Volumen del <i>dataset</i> sin respuestas y su tiempo de procesamiento.	27
2	Volumen del <i>dataset</i> con respuestas y su tiempo de procesamiento.	29
3	Resultado del análisis de sentimiento aplicado al <i>dataset</i> con respuestas	30
4	Positividad, neutralidad y negatividad presente en el <i>dataset</i> con respuestas . .	31
5	10 <i>Hashtags</i> más recurrentes encontrados en el <i>dataset</i>	32
6	10 <i>Hashtags</i> más recurrentes encontrados en el <i>dataset</i> , clasificados por sentimiento.	33
7	7 usuarios que más <i>tweets</i> publicaron sobre la convención constituyente entre enero y septiembre del 2022.	35
8	6 usuarios que más <i>tweets</i> publicaron sobre la convención constituyente entre enero y septiembre del 2022, categorizados por sentimiento.	36
9	10 usuarios más mencionados en <i>tweets</i> sobre la convención constituyente entre enero y septiembre del 2022.	37
10	10 constituyentes más mencionados en <i>tweets</i> sobre la convención constituyente entre enero y septiembre del 2022.	38
11	10 constituyentes más mencionados en <i>tweets</i> sobre la convención constituyente entre enero y septiembre del 2022, categorizados por sentimiento.	40
12	10 no constituyentes más mencionados en <i>tweets</i> sobre la convención constituyente entre enero y septiembre del 2022.	41
13	10 no constituyentes más mencionados en <i>tweets</i> sobre la convención constituyente entre enero y septiembre del 2022, categorizados por sentimiento.	44
14	10 medios de comunicación más mencionados en <i>tweets</i> sobre la convención constituyente entre enero y septiembre del 2022.	45
15	10 medios más mencionados en <i>tweets</i> sobre la convención constituyente entre enero y septiembre del 2022, categorizados por sentimiento.	46

16	Volumen del <i>dataset</i> con respuestas y su tiempo de procesamiento para análisis de emociones.	48
17	Análisis de emociones aplicado sobre el <i>dataset</i> con respuestas	49
18	10 <i>Hashtags</i> más recurrentes encontrados en la clasificación por emoción.	50
19	Detección de discurso de odio aplicada sobre el <i>dataset</i> con respuestas	51
20	Detección de discurso de odio en <i>tweets</i> referentes a Chile plurinacional e intercultural	53
21	Detección del aumento en discurso de odio provocado por <i>tweets</i> referentes a Chile plurinacional e intercultural	53
22	10 <i>Hashtags</i> más recurrentes encontrados en la detección de discurso de odio.	55
23	Respuestas del grupo de validación en el cuestionario de entrada, segmentado por sexo.	84
24	Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 9, segmentado por sexo.	85
25	Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 11, segmentado por sexo.	86
26	Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 16, segmentado por sexo.	86
27	Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 17, segmentado por sexo.	87
28	Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 20, segmentado por sexo.	87
29	Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 28, segmentado por sexo.	88
30	Porcentaje de éxito y fracaso obtenido por el grupo de validación	89
31	Respuestas del grupo de validación en el cuestionario de salida, segmentado por sexo.	90

INTRODUCCIÓN

Según la Real Academia Española, la Constitución se define como “la ley fundamental de un Estado, con rango superior al resto de las leyes, que define el régimen de los derechos y libertades de los ciudadanos y delimita los poderes e instituciones de la organización política” [Asale y Rae, 2020], es decir, la norma jurídica más importante del Estado, bajo la cual se rigen todos los ciudadanos del país, estableciendo la regulación política y administrativa, instaurando la separación de los poderes del Estado, reconociendo la soberanía nacional y los derechos fundamentales de las personas.

En Chile, la última Constitución data de 1980 y define como uno de los métodos para reescribirla el plebiscito nacional, una votación en la cual todos los ciudadanos con derecho a sufragio pueden participar, para decidir si se desea iniciar un proceso constitucional, el cual bajo normas previamente definidas podrá determinar el mecanismo que estará a cargo de redactar una nueva Carta Magna.

Con el crecimiento exponencial de las redes sociales en los últimos años y la irrupción de los medios digitales, el debate se ha llevado a una plataforma virtual y más democrática, en la que cualquier persona puede vertir su opinión libremente independiente de su condición socioeconómica o de sus estudios académicos, entregando así la oportunidad de conocer las opiniones de un número mucho más grande de personas, lo cual antes de la era de las redes sociales era impensable.

Tomando el punto cúlmine del reciente proceso constitucional, el plebiscito constitucional de salida del 2022, la presente memoria buscará identificar aquellos hechos y tendencias que marcaron su resultado mediante el periodismo de datos, recabando noticias acerca de los sucesos claves ocurridos entre inicios del 2022 hasta la fecha en que se llevó a cabo el plebiscito, así como también extrayendo opiniones de los votantes desde la red social Twitter, con el fin de identificar su percepción respecto al proceso y a diferentes tópicos dentro del mismo.

La estructura de esta memoria consta de cinco capítulos. El *Capítulo 1: Definición del problema* inicia contextualizando el escenario electoral en nuestro país entre los años 2020 y 2022, para luego describir los objetivos de este trabajo y finalmente definir su alcance. El siguiente capítulo, *Marco Conceptual*, hace un repaso por las principales técnicas que serán empleadas en esta memoria, definiéndolas y explicando su utilidad para la consecución de los objetivos descritos en el capítulo anterior. De la misma manera, también se estudian trabajos anteriores relacionados a las tecnologías a usar, con el fin de que sirvan de base para la metodología que se desea implementar. Luego, en el Capítulo 3, se propone una solución al problema previamente descrito, mediante la elección de una metodología, describiendo su ejecución paso a paso, para finalmente llegar a la validación de la propuesta, donde se verificará si lo entregado por el estudio cumple sus objetivos, obteniendo y analizando sus resultados, con el fin de entregar conclusiones en el último capítulo.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

El primer capítulo de esta memoria trata acerca de la definición del problema, donde se introduce brevemente el escenario electoral en nuestro país durante los últimos años, para luego definir los objetivos generales y específicos que fueron conseguidos, junto con el alcance que este trabajo puede otorgar para futuras investigaciones o para la sociedad chilena en general.

1.1. Escenario electoral de Chile entre el 2020 y el 2022

El plebiscito constitucional de salida de Chile de 2022 fue el punto cúlmine del proceso constitucional que vivió nuestro país entre los años 2020 y 2022. En él, la ciudadanía debió aprobar o rechazar la propuesta elaborada por la Convención Constitucional que fue entregada al presidente de la República, Gabriel Boric, el 4 de julio de 2022, para que fuese votada el día 4 de septiembre del mismo año, según lo acordado entre la mesa directiva de la Convención Constitucional y la Secretaría General de la República.

En este mismo lapso, Chile vivió una cantidad considerable de procesos electorarios, los cuales marcaban un antecedente para lo que iba a ser el plebiscito de salida. Entre dichos procesos, destacan el plebiscito de entrada de 2020 y la elección presidencial del 2021. El primero, producto del acuerdo constitucional del 2019 [Senado, 2019], liderado por los partidos políticos para descomprimir el tenso ambiente sociopolítico de ese entonces, hizo crecer la idea de que el país necesitaba cambios profundos, tendencia que se arrastró hasta el resultado del plebiscito, donde la opción Apruebo (una nueva constitución) obtuvo un triunfo categórico con el 78,28 % de las preferencias, eligiendo como mecanismo para redactar ésta a la Convención Constitucional, con un 79 % de los votos.

Un año más tarde, en medio del proceso constitucional, se llevaron a cabo las elecciones presidenciales del 2021, que para la sociedad chilena eran vistas como una prueba para ver si el escenario político realmente había cambiado luego del contundente triunfo del Apruebo. En un inicio, no parecía ser así, luego del sorprendente triunfo en primera vuelta de José Antonio Kast, abanderado del partido de extrema derecha, Republicanos, con un 27,91 % de las preferencias, superando a figuras como Gabriel Boric (Apruebo Dignidad, izquierda) y Sebastián Sichel (Chile Vamos, centroderecha). Sin embargo, en segunda vuelta la tendencia se revertiría, siendo finalmente electo Gabriel Boric como el presidente de la República, obteniendo el 55,87 % de los votos.

Con estos antecedentes, se esperaba un resultado similar para el plebiscito constitucional de salida, con Gabriel Boric en el poder, siendo uno de los principales impulsores del acuerdo constitucional del 2019 y perteneciente a Apruebo Dignidad, la coalición que más activa-

mente hizo campaña por el Apruebo, tanto en el 2020 como en el 2022. Sin embargo, con el paso del tiempo la Convención Constitucional fue perdiendo respaldo debido a noticias, declaraciones, rumores y especulaciones que fueron sembrando dudas en la ciudadanía, la cual lentamente comenzó a cambiar de opinión e inclinarse hacia la opción del Rechazo.

A fines de enero del 2022, la encuesta Cadem marcaba un apoyo del 56 % al Apruebo, una tendencia similar a la vista en la elección presidencial del año anterior, si se hace la comparación con el porcentaje de votos obtenidos por Gabriel Boric. A medida que comenzaron a circular las noticias y declaraciones mencionadas en el párrafo anterior, la distancia entre ambas opciones fue disminuyendo, al punto de revertirse la tendencia a comienzos de abril, la última vez que el Apruebo lideró las encuestas, con ya un solo 39 % de las preferencias, es decir, una baja de 17 puntos porcentuales en un poco más de dos meses. De este punto en adelante, el Rechazo arrasó con todas las encuestas, posicionándose con una amplia ventaja que llegó a 16 puntos en la encuesta de AtlasIntel, registrada entre el 16 y 19 de agosto, a casi dos semanas del plebiscito. Finalmente, el Rechazo fue proclamada la opción vencedora, con un sorprendente 61,89 % contra un 38,11 % del Apruebo [BCN, 2022], una diferencia que ninguna encuesta logró anticipar, lo cual se puede visualizar en la Figura 1.

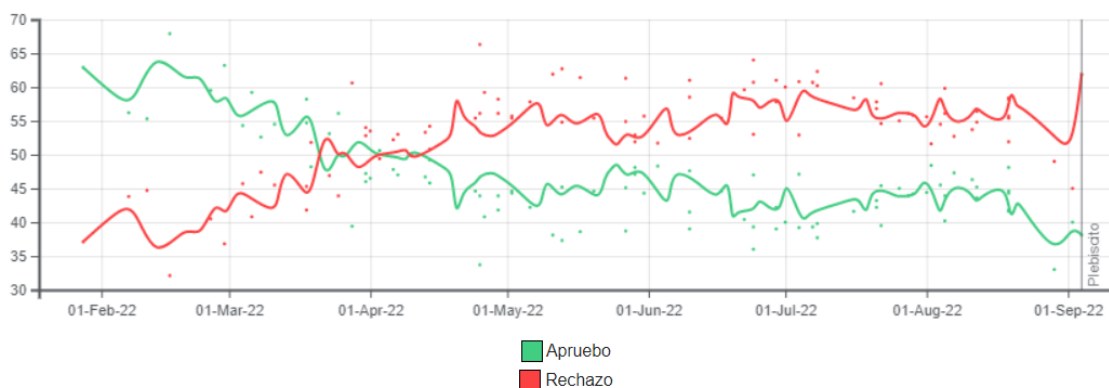


Figura 1: Evolución de las encuestas del plebiscito constitucional de Chile de 2022.
Fuente: Wikipedia (basado en los resultados oficiales entregados por las agencias encuestadoras).

Múltiples razones pueden explicar el declive del Apruebo, tanto en hechos circunstanciales como coyunturales, los cuales han sido expuestos y analizados por políticos, historiadores e incluso los mismos protagonistas del proceso. Sin embargo, desde el punto de vista de los datos, ¿qué análisis podemos hacer? ¿es posible usar la informática para ayudar a hallar las causas que llevaron a este resultado?

1.2. Objetivo de la solución

El objetivo de este trabajo es emplear el periodismo de datos para identificar aquellos momentos claves en los cuales las tendencias sufrieron sus mayores variaciones, y así conocer más acerca del debate que terminó llevando al triunfo del Rechazo. Mediante la recopilación de noticias provenientes desde medios de comunicación no tradicionales en el periodo comprendido entre enero y agosto de 2022, y la obtención de declaraciones y opiniones en la red social Twitter utilizando *web scraping*.

1.2.1. Objetivo general

Identificar tendencias y hechos que más influenciaron el resultado del Plebiscito Constitucional de Chile en el año 2022, mediante el periodismo de datos, para aportar, desde la informática, con más información relevante que permita profundizar los estudios y análisis ya existentes.

1.2.2. Objetivos específicos

- Recopilar noticias e información sobre el plebiscito en un lapso de 6 meses previo a la votación para su posterior procesamiento y análisis, con el objetivo de identificar los hechos trascendentales ocurridos durante el proceso constitucional que puedan significar puntos de inflexión en las tendencias.
- Diseñar un algoritmo de procesamiento de texto para identificar los principales temas que predominaron el debate constitucional en Twitter, con el fin de generar un análisis en base a los tópicos recurrentes en la discusión.
- Emplear análisis de sentimiento para analizar la percepción de los votantes respecto a determinados tópicos, para generar una imagen de la percepción de los usuarios de Twitter respecto a la convención constituyente y su trabajo.
- Diseñar una metodología que permita determinar el grado de influencia de un usuario, y así identificar a aquellos personajes que lograron influenciar a una gran cantidad de personas.

CAPÍTULO 2

MARCO CONCEPTUAL

El presente capítulo tiene como objetivo definir desde un punto de vista conceptual la metodología, técnicas y herramientas utilizadas para resolver el problema definido en el capítulo anterior, cuyo uso será especificado en el *Capítulo 3: Propuesta de Solución*, y profundizado en el *Capítulo 4: Validación de la solución*.

2.1. Periodismo de datos

El periodismo de datos, según Sandra Crucianelli [Crucianelli, 2013] es una mezcla entre cinco diferentes disciplinas del periodismo, entre las cuales se encuentran:

- **Periodismo de investigación:** Hace referencia al periodismo hecho sobre un tema de relevancia social, como el proceso constitucional de Chile, a lo que se añade también el hecho de que alguien quiera mantenerlo en secreto, en este caso, las tendencias y rumores en redes sociales sobre dicho proceso que aún no han sido estudiadas.
- **Periodismo de profundidad:** Lo mismo que el periodismo de investigación, pero en este caso, nadie intenta mantener un secreto.
- **Periodismo de precisión:** Periodismo que involucra el uso de métodos de la investigación social aplicados al periodismo, como por ejemplo la investigación cualitativa, de lo cual se desprende el análisis del discurso.
- **Periodismo analítico:** Se refiere al área del periodismo que se dedica a analizar realidades complejas, con el fin de lograr la comprensión del público, mediante la descripción de detalles históricos y datos estadísticos.
- **Periodismo asistido por computadora:** Es definido como cualquier proceso que utiliza la asistencia de computadoras durante la recolección y procesamiento de datos.

Junto a esto, se incorporan conceptos propios de la informática, como lo son:

- **Grandes volúmenes de datos:** Conocido también como el *big data*, se refiere a grandes colecciones de datos semiestructurados o no estructurados, las cuales pueden ser empleadas para futuros estudios y/o análisis.
- **Visualización interactiva:** O visualización de datos, se encarga de mostrar información en un formato más cómodo para el lector, con el objetivo de hacer más clara la comprensión de una determinada idea.

- **Programación:** La base de la recopilación y análisis de datos, junto con la visualización de éstos, se basa en diseñar y poner a prueba algoritmos que permitan agilizar dichos procesos, mediante la automatización de éstos.

Incorporar la informática al mundo del periodismo de datos abre una nueva ventana para tener panoramas más completos sobre sucesos sociales, a través del procesamiento de fuentes de datos con una capacidad que en tiempos anteriores no era imaginable.

2.2. Metodologías para el análisis de datos

Dado que el periodismo de datos requiere analizar los mismos, es necesario definir una metodología de trabajo que sea consistente con el propósito deseado, la naturaleza de los datos trabajados, y las metodologías empleadas. A continuación, se detallan tres metodologías para el análisis de datos que son ampliamente usadas y podrían ser adoptadas para el desarrollo de esta investigación:

- **CRISP-DM (*Cross-industry standard process for data mining*):** Es una metodología que separa el proceso de análisis de datos en 6 fases, las cuales contemplan el entendimiento del negocio, el entendimiento de los datos, el preprocesamiento de los datos, el modelado, la evaluación y finalmente la implementación, las cuales siguen una naturaleza lineal con pasos que no necesariamente son secuenciales, dado que el entendimiento y el preprocesamiento pueden ser subprocesos iterativos, así como la evaluación puede determinar que se requiere un mayor entendimiento del negocio antes de continuar con la implementación, es decir, volver al primer paso.
- **SEMMA (*Sample, Explore, Modify, Model, and Assess*):** Es un proceso secuencial utilizado ampliamente en la inteligencia de negocios y la estadística. El primer paso *Sample*, se refiere a la selección de los datos para el modelado, que deben ser suficientes para el propósito que se desea lograr; luego, *Explore* se refiere a un análisis exploratorio, con el fin de detectar correlaciones inesperadas entre las variables del *dataset* definido en el paso anterior, con la ayuda de la visualización de datos. *Modify* se refiere a la transformación de los datos en caso de ser necesario, para posteriormente pasar a modelar (*Model*), aplicando análisis de datos para obtener los resultados deseados, que finalmente serán evaluados en la etapa de *Assess*, probando la confiabilidad de los modelos diseñados.
- **KDD (*Knowledge Discovery in Databases*):** Es un proceso secuencial simple de 5 etapas enfocado en la extracción automatizada de datos que puede repetirse las veces necesarias para obtener los resultados esperados, teniendo también así, un carácter iterativo. KDD inicia con la recopilación de los datos, para su posterior selección, limpieza y transformación, que permitirán la aplicación del análisis de datos, mediante el uso de modelos que deberán ser evaluados en la etapa siguiente, logrando así, en el

paso final, la interpretación de los datos y las deducciones correspondientes en base a éstos.

2.3. *Web scraping*

Web scraping es el proceso de recolectar datos contenidos en páginas web mediante técnicas automatizadas, las cuales deben ser ideadas mediante un algoritmo para luego ponerlas en práctica, identificando la estructura de la información que se desea recuperar, dado que en la web los datos mayoritariamente corresponden al tipo de no estructurados, por ende, el objetivo es identificar el patrón que siguen estos, capturando únicamente los datos necesarios para los propósitos que se estimen convenientes.

El *web scraping* en Twitter, supone un caso especial, pero bastante común, dado que existen múltiples trabajos similares, pero sobre diferentes tópicos, incluso dentro del contexto de las memorias del Departamento de Informática de la UTFSM. Se dice que es un caso especial dado que Twitter cuenta con herramientas especializadas para desarrolladores con el objetivo de facilitar este trabajo y abrir la información hacia el público general. Dichas herramientas especializadas son las APIs REST y Streaming de Twitter, cuyo funcionamiento se encuentra detallado en el trabajo de Espin, Cruz y Gamboa, titulado “Métodos de extracción de comentarios de la red social Twitter para uso en Procesamiento de Lenguaje Natural” [Espin-Riofrío *et al.*, 2021]. La API REST recolecta *tweets* publicados recientemente, mientras que la API Streaming recolecta *tweets* publicados en el momento, por lo cual esta última no será utilizada, dado que la información que se desea extraer ya existe. Python también cumple un rol fundamental en el uso de estas APIs, ya que puede apoyar mediante el uso de bibliotecas adicionales al funcionamiento del proceso, como Tweepy y Twint.

Sin embargo, durante el tiempo de confección de este documento, Twitter sufrió severos cambios en su estructura y la forma de extraer información desde ésta. La API de Twitter fue reemplazada por su segunda versión que ya no incluye beneficios ni facilidades para aquellas personas que buscan hacer investigación en base a la información que puede encontrarse en esta red social. Así, surgen otras alternativas como Snsrape, una biblioteca de *web scraping* multipropósito con módulos especializados para la extracción masiva de *tweets*, cuya estructura se detalla en la Figura 2, que además permiten extraer métricas claves para su estudio, como cantidad de *likes*, respuestas, *retweets*, entre otros.

Existen también herramientas adicionales que permiten hacer *web scraping* tanto manual como automatizado, como es el caso de Octoparse, que con sólo ingresar una URL puede identificar su estructura automáticamente y extraer toda la información disponible, mientras que el método manual permite al desarrollador seleccionar los datos específicos que se desean obtener, lo cual en el contexto de este trabajo puede resultar bastante útil.

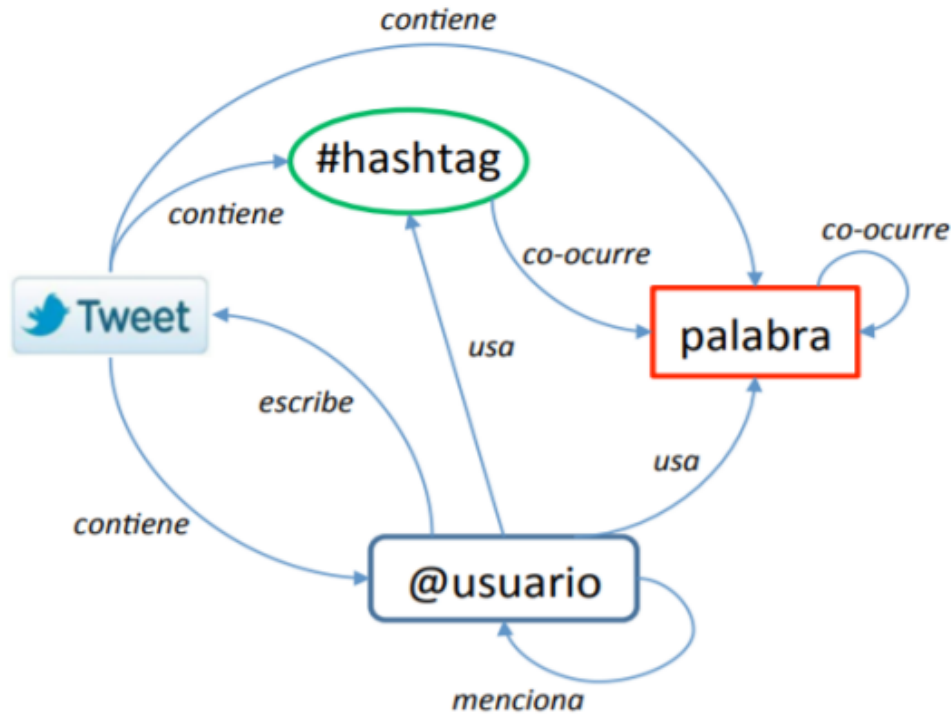


Figura 2: Anatomía de un tweet.
Fuente: [Espín-Riofrío *et al.*, 2021].

2.4. Análisis de sentimiento

Aquí es donde entra el análisis de sentimiento o minería de opinión, una rama del Procesamiento del Lenguaje Natural, que es el estudio computacional de los sentimientos, emociones y opiniones expresadas en texto, la cual perfectamente puede ser aplicada para las opiniones emitidas a través de redes sociales como Twitter, razón por la cual ya existe documentación y trabajos a la fecha, como el de Sarlan, Nadam y Basri en 2014, titulado “*Twitter Sentiment Analysis*” [Sarlan y Basri, 2014] donde se extraen *tweets* para luego analizar si las opiniones encontradas son positivas, negativas o neutras respecto a determinado tópico, utilizando las APIs de Twitter, Python, Django y otras herramientas como *machine learning*, que a diferencia de las primeras técnicas utilizadas en análisis de sentimiento como el *lexicon-based approach*, basándose en “tokenizar” los tweets en base a un diccionario base que contiene información clave sobre la opinión que se desea analizar, iniciándose así un “punteo de polarización”, el cual debe superar un *threshold* para ser considerada positiva, o en caso contrario, negativa.

El mismo trabajo nombra las ventajas de realizar análisis de sentimiento basado en *machine learning* cuando se trata de Twitter, dado que es posible crear, entrenar y adaptar modelos para propósitos y contextos específicos. A diferencia del método anterior, se identifican todos

los adjetivos presentes en un *tweet*, para luego crear un *word set* con los N adjetivos más populares; finalmente analizando cada *tweet* extraído en base a tres criterios: número de palabras positivas, número de palabras negativas y la frecuencia de cada palabra, de acuerdo con el modelo entrenado que se haya diseñado para el propósito específico, será clasificado finalmente como una opinión positiva, negativa o neutra sobre un tema.

Un año más tarde, Sahayak, Shete y Pathan [Sahayak *et al.*, 2015], ahondarían más en los algoritmos de *machine learning* que pueden ser empleados para tareas de clasificación de texto, o en este caso, análisis de sentimiento. Si bien a la fecha existen múltiples métodos que pueden ser utilizados para esta misma tarea, en esta sección se profundizarán los tres algoritmos más populares y exitosos: *Naive Bayes*, *Máxima Entropía* y *Support Vector Machines* (SVM).

- *Naive Bayes*: O Bayes “ingenuo” en español, es un algoritmo en el cual los clasificadores de texto asumen que el valor de una variable en una determinada clase es independiente del efecto de otras variables, es decir, no se ve afectado por el valor de otras variables. A esto se le llama independencia condicional, y se extrae directamente del Teorema de Bayes y la definición de la probabilidad condicional, lo cual puede ser aplicado al análisis de sentimiento en *tweets* contando la cantidad de *features* específicas presentes en éste, es decir, alguna palabra en específico como un adjetivo, verbo, adverbio, etc, con el objetivo de identificar si la percepción del mensaje es positiva, negativa o neutral respecto a un determinado tema.

De acuerdo a la frecuencia de aparición de ciertas palabras, un determinado *tweet* puede pertenecer a una determinada clase, lo cual mediante inferencia bayesiana se puede calcular una probabilidad, mediante la siguiente fórmula:

$$c = \operatorname{argmax} P(x | d); x \in C, d \in D$$

Sin embargo, teniendo en cuenta el Teorema de Bayes, cuya fórmula es:

$$P(x | y) = \frac{P(y | x)P(x)}{P(y)}$$

se puede finalmente expresar la probabilidad c calculada anteriormente como:

$$c = \operatorname{argmax} \frac{P(d | x)P(x)}{P(d)}; x \in C, d \in D$$

Obteniendo así, una forma de clasificar los *tweets* en base a las palabras más recurrentes que aparezcan en éste.

- *Máxima entropía*: A diferencia de *Naive Bayes*, este algoritmo se basa en características o *features*, en donde no se hacen suposiciones respecto de la independencia

de éstas, haciendo que para este algoritmo sea más fácil manejar problemas de solapamiento gracias al uso de un *Stanford classifier* [Stanford, 2012], que categoriza los datos entregados en conjuntos, a los cuales se les asigna un peso positivo o negativo, que pueden ser asignados manualmente en un inicio, pero que gradualmente deberían ser “aprendidos” por el modelo mediante el entrenamiento de éste, con el fin de que el peso que se le entregue a las palabras dependa del contexto en el que se presenta, ya que podría adoptar otro significado en un escenario diferente y tener lecturas erróneas acerca de si la opinión vertida es positiva, negativa o neutral.

Para este algoritmo, la idea no es calcular mediante el teorema de Bayes tal como se presenta, sino que se desea calcular el $P(y | x)$, el cual se da de la siguiente manera:

$$P(y | x) = \sum_{i=1}^{\infty} w_i f_i$$

con w_i y f_i correspondiendo a las características o *features* de cada *tweet* y su respectivo peso. Sin embargo, dicha expresión requiere de discretización para la tarea que se desea emplear, es decir, obtener resultados que se encuentren entre 0 y 1, para lo cual se propone la siguiente fórmula:

$$p(y = c | x) = \frac{1}{Z} \exp \left(\sum_i w_i f_i(c, x) \right)$$

siendo Z el denominador que discretiza los resultados de la ecuación, el cual puede ser reemplazado por una variable N , que corresponde al número de características de cada *tweet*, para obtener la probabilidad de que y pertenezca a la clasificación c dado x :

$$p(c | x) = \frac{\exp \left(\sum_{i=1}^N w_i f_i(c, x) \right)}{\sum_{d \in C} \exp \left(\sum_{i=1}^N w_i f_i(d, x) \right)}$$

- **Support Vector Machines (SVM):** Desarrollados en el ámbito de la teoría del aprendizaje estadístico, las SVM son algoritmos compuestos de diversos métodos de aprendizaje que pueden emplearse tanto para clasificación como para regresión, por lo cual se utilizan, principalmente, en reconocimiento de patrones. También se basan en la dimensión de Vapnik-Chervonenkis, que recibe el nombre de sus autores, Vladimir Vapnik y Alexey Chervonenkis, la cual es una medida de capacidad de los algoritmos de clasificación estadística, más específicamente la cardinalidad del mayor conjunto de puntos separables por un determinado método. Para *machine learning*, la dimensión de Vapnik-Chervonenkis se emplea para predecir correctamente el límite superior probabilístico sobre el error en un conjunto de prueba o *testeo*, logrando que el entrenamiento del modelo sea un proceso más pulcro y libre de errores, obteniendo así

mejores resultados al momento de utilizar el algoritmo para análisis de sentimiento en base a un *tweet*, dependiendo del tema sobre el cual queremos medir su percepción. Las SVM intentan buscar un hiperplano que separe los datos en dos clases de la forma más óptima posible, como se ejemplifica en la Figura 3.

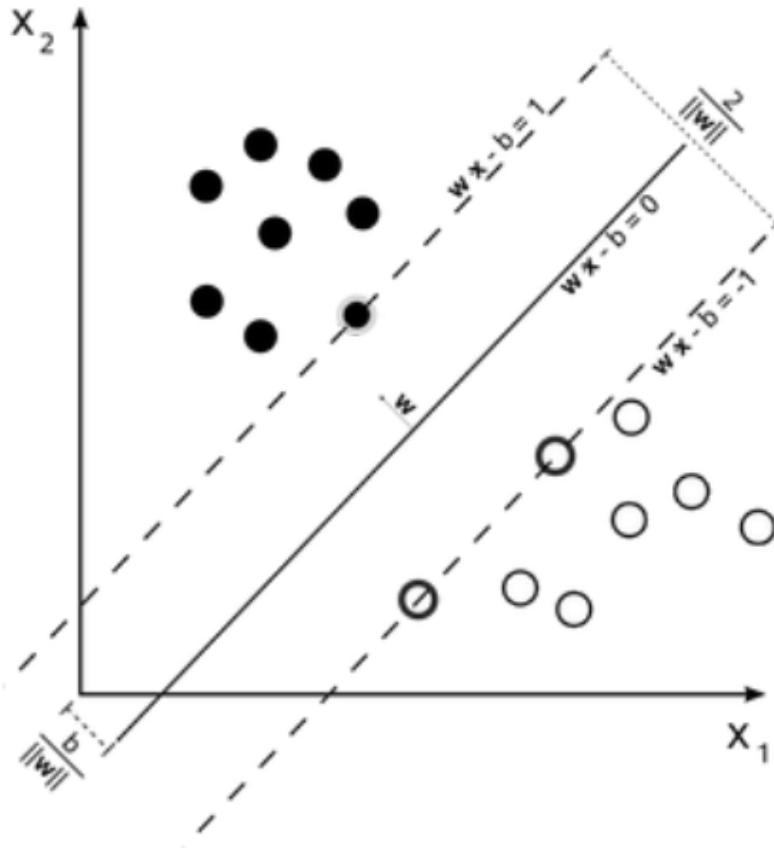


Figura 3: Representación de un hiperplano en una SVM.
 Fuente: Elaboración propia (basado en [Sahayak *et al.*, 2015]).

con el hiperplano representado por la siguiente ecuación:

$$\langle \vec{w} \vec{x} \rangle + b = \sum_i y_i \alpha_i \langle \vec{x}_i \vec{x} \rangle + b = 0$$

Así, el proceso descrito por Sahayak, Shete y Pathan se vería según lo descrito en la Figura 4.

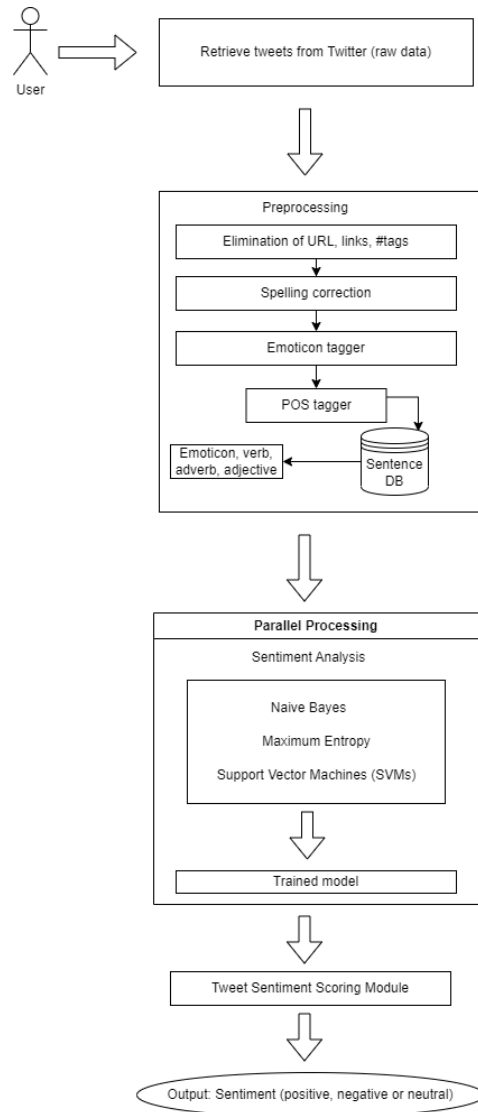


Figura 4: Estructura del análisis de sentimiento para *tweets* usando *machine learning*.
Fuente: Elaboración propia (basado en [Sahayak et al., 2015]).

En base a estos algoritmos, es posible tener un primer acercamiento a un algoritmo que permita extraer *tweets* o noticias de medios de comunicación no tradicionales para luego ser sometidas al análisis del sentimiento. La estructura de dicho algoritmo, constaría de los siguientes cinco pasos:

1. Extracción de información: Mediante el uso de *web scraping*, que ya fue detallado anteriormente, se crearán dos *data sets* diferentes, uno compuesto de noticias y el otro de *tweets*, los cuales serán trabajados bajo la misma metodología pero de manera independiente, con el fin de no combinar ambas fuentes de información, dado que tienen estructuras e intenciones diferentes, lo cual no podría ser soportado por un

único modelo de *machine learning*.

2. Preprocesamiento: Una vez extraída la información, se debe aplicar un preprocesamiento a los datos, filtrando y eliminando palabras, *tokenizando* y construyendo *n-grams*, dado que si se entrena y se aplica el modelo a la información tal y como está, los resultados y el rendimiento no serían óptimos, lo cual puede llevar a conclusiones incorrectas acerca de las intenciones y sentimientos vertidas en las opiniones analizadas.
 - a) Filtrado: En esta etapa se eliminan las URL, palabras reservadas de Twitter como RT o *retweet*, emoticones y menciones a usuarios, como por ejemplo @usuario123.
 - b) *Tokenización*: Como ya fue explicado previamente dentro de esta misma sección, *tokenizar* se refiere a la acción de segmentar oraciones, formando conjuntos de palabras al eliminar espacios y signos de puntuación.
 - c) Borrado de “palabras vacías” o *stopwords*: Aquí se remueven palabras como artículos (el, la, los, las, etc.) o conectores y preposiciones como a, o, y, entre otros, dado que no aportan contexto al contenido del mensaje y dificultan el proceso de entrenamiento del modelo.
 - d) Construcción de *n-gramas*: Finalmente, se construyen *n-gramas* como conjuntos de palabras consecutivas en los cuales se debe tener especial atención a la negación, dado que la existencia de ésta cambia completamente la polaridad de dicho segmento; por ende, también es importante considerar que palabras como “no”, podría no corresponder a la categoría de palabra vacía dependiendo del contexto, que deberá ser analizado por el modelo entrenado.
3. Entrenamiento: Para entrenar al modelo de *machine learning* se debe pasar por un proceso de clasificación y predicción y que se complementan entre sí, dado que el clasificar asigna etiquetas, pesos y polaridades para diferentes conjuntos de palabras, con el objetivo de que la predicción se encargue de determinar a qué clase corresponden aquellos segmentos que no han sido clasificados. Es en este paso donde se construye un modelo a partir del entrenamiento y se utilizan los algoritmos mencionados anteriormente, como *Naive Bayes* y *Support Vector Machines*.
4. Puntuación del sentimiento: Con el modelo ya entrenado, es posible determinar la polaridad de las palabras, asignando puntajes que van de lo negativo a lo positivo, pasando por el neutro.
5. Clasificación del sentimiento: Una vez los puntajes ya están determinados, se calcula la polaridad de la noticia o el *tweet* completo para determinar si corresponde a una opinión positiva, negativa o neutra respecto a un determinado tema.

2.5. Métodos de validación para algoritmos de análisis de sentimiento basados en *machine learning*

Dado que el *machine learning*, como una rama de la inteligencia artificial, puede cometer errores, es importante contar con métodos que permitan determinar la precisión y exactitud de los modelos utilizados, con el fin de asegurar que los resultados obtenidos son confiables y permiten extraer conclusiones verídicas. A continuación se introducen dos métricas comúnmente utilizadas para la validación de modelos basados en *machine learning*.

- Matriz de confusión: Es una matriz de 2x2, en el caso de tener una salida binaria, que cuenta los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, entregando mediante métricas como la precisión y la exactitud la calidad de un modelo en *machine learning*, así como también de cualquier algoritmo que emplee Inteligencia Artificial. Los verdaderos positivos (*VP*) y negativos (*VN*) corresponden a predicciones correctas, mientras que los falsos positivos (*FP*) y negativos (*FN*) a predicciones incorrectas. La precisión de un modelo se calcula mediante la ecuación:

$$precision = \frac{VP}{VP + FP}$$

mientras que la exactitud se define a través de la siguiente fórmula:

$$exactitud = \frac{VP + VN}{VP + FP + VN + FN}$$

Valor Verdadero	Sí	VP	FP
	No	FN	VN
		Sí	No
		Valor Predicho	

Figura 5: Ejemplo de una matriz de confusión.
Fuente: Elaboración propia.

Se dice que un buen modelo debe ser exacto y preciso a la vez, es decir, que sea capaz de tener una alta tasa de reconocimiento efectiva para verdaderos positivos y negativos, pero que también sea bueno en detectar los falsos positivos y negativos. En cuanto a números, un modelo se considera aceptable dependiendo de sus características y espacio muestral, por lo que los porcentajes deben oscilar entre un 70 %-80 %.

- Curva ROC (*Receiver Operating Characteristic*): Es una representación gráfica de la sensibilidad frente a las falsas alarmas de un sistema de clasificación binario. Dicha curva corresponde a la comparación entre la proporción de verdaderos positivos, VPR , versus la proporción de falsos positivos, FPR , las cuales se calculan de la siguiente manera:

$$VPR = \frac{VP}{P} = \frac{VP}{VP + FN}$$

$$FPR = \frac{FP}{N} = \frac{FP}{FP + VN}$$

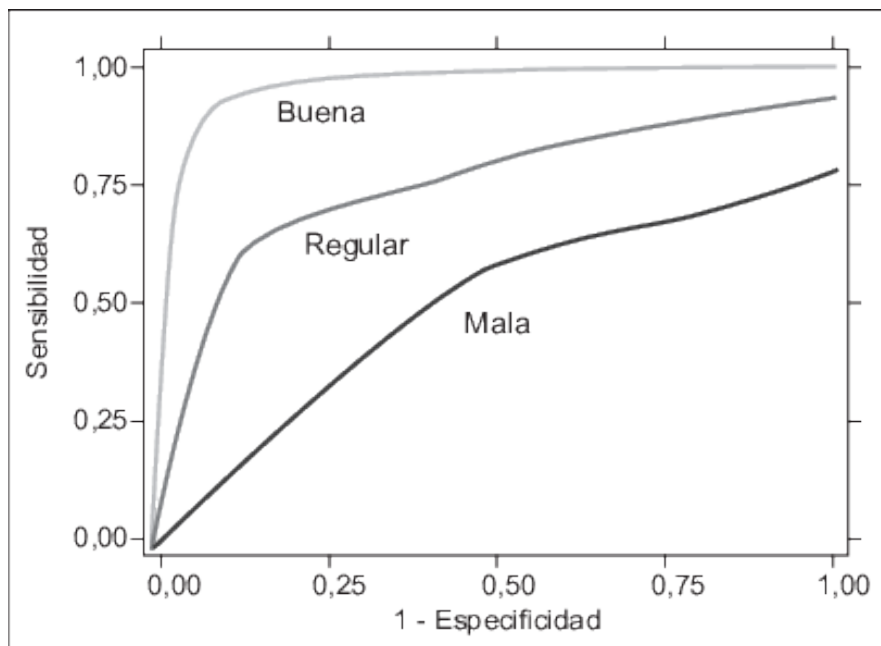


Figura 6: Ejemplo de una Curva ROC.

Fuente: ResearchGate.

Como se ve en la Figura 6, una curva ROC es mejor a medida que aumenta el área bajo la curva (AUC), cuyo valor puede oscilar entre 0,5 y 1. Un mal modelo marca valores entre 0,5 y 0,75, mientras que un buen modelo lo hace entre 0,75 y 0,9. Sobre 0,9 ya se está hablando de un modelo excepcional en cuanto a rendimiento, ya que se encuentra cerca de la perfección teórica.

2.6. Visualización de datos

La información visual desempeña un papel fundamental en la comprensión de nuestro entorno y experiencias. Los colores, las formas y los patrones son elementos cruciales para transmitir y recibir información en diversos medios, como la escritura, las imágenes, los videos y, por supuesto, las visualizaciones de datos.

Las visualizaciones de datos son una herramienta esencial en este contexto, ya que permiten presentar información de manera eficaz y fácilmente comprensible. Por lo tanto, es fundamental comprender cómo diseñar visualizaciones efectivas y qué elementos deben incluirse para lograrlo. A continuación, se detallan los aspectos clave que deben considerarse al crear visualizaciones de datos y se describe la metodología para llevar a cabo este proceso.

Es importante tener en cuenta es el "lenguaje visual". Este lenguaje abarca diversas dimensiones que deben ser coherentes entre sí para facilitar la comprensión del espectador y contribuir a la narrativa visual. Estas dimensiones del lenguaje visual en las visualizaciones de datos pueden incluir:

- **Color:** Los colores desempeñan un papel crucial en la representación de datos y constituyen un aspecto importante del lenguaje visual. La elección de colores puede ayudar a distinguir categorías o escalas de valores, ya sea mediante la asignación de colores diferentes a cada categoría o mediante la variación de tonalidades dentro de un mismo color para representar rangos. Además, ciertos colores destacados pueden utilizarse para resaltar información relevante en la visualización.
- **Geometría de las formas:** Las visualizaciones pueden adoptar diferentes formas geométricas, las cuales pueden tener significados relacionados con la categorización de los datos o representar elementos visuales como gráficos y mapas. Es esencial definir el significado de estas formas si se utilizan como parte del lenguaje visual.
- **Estructura:** Es importante definir qué fenómeno, estructura o proceso se pretende visualizar con la visualización de datos. Las visualizaciones pueden adoptar diversas formas, como descriptivas, secuenciales, narrativas, flujos o relaciones, según la naturaleza de los datos y el objetivo de la representación.
- **Tipografía:** La elección de la tipografía también influye en el lenguaje visual de las visualizaciones. Diferentes tipos de fuentes pueden transmitir mensajes distintos, como el uso de fuentes con serifas para asociar los datos a textos impresos o fuentes de estilo más creativo para representar datos informales o menos formales.

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

Luego de definir de forma conceptual la metodología, algoritmos, herramientas y técnicas sobre los cuales se apoyó el trabajo descrito en este documento, a continuación se detalla la propuesta de solución que se llevó a cabo para resolver el problema descrito en el Capítulo 1, en el cual se llevan a un plano más práctico los conceptos previamente definidos, detallando para qué fueron empleados y cuál fue su rol en la resolución de la problemática. Dicha propuesta comienza en la elección de la metodología que fue empleada para llevar a cabo este trabajo, KDD, para luego detallar las herramientas y técnicas que se usaron en cada etapa.

3.1. Elección de la metodología

En el marco conceptual se habla acerca de tres metodologías que pueden ser utilizadas para proponer la solución que se llevó a cabo en este estudio, CRISP-DM, SEMMA y KDD, de las cuales finalmente una fue la elegida. El criterio que se utilizó para la selección de la metodología se basó en la disponibilidad de las herramientas que se emplearon y el orden lógico que se les deseó dar. CRISP-DM es una metodología muy aplicada a negocios, la cual no resulta ser idónea para el problema previamente definido, que es de carácter investigativo. Por su parte, SEMMA presenta una estructura similar al trabajo que se realizó, pero considera el uso de visualizaciones de datos previo al procesamiento y modelado, lo cual no conecta con el propósito de utilizar la visualización de datos para presentar los resultados obtenidos luego de dichos pasos. Finalmente, KDD es un proceso simple que se ajusta a los objetivos de esta investigación, mediante la extracción automatizada de datos, su selección, limpieza y transformación, posterior modelado e interpretación de los resultados, donde se pueden emplear las visualizaciones de una forma que cronológicamente hacen sentido para este estudio. Por ende, KDD fue seleccionada como la metodología que se empleó a lo largo de todo el proceso.

El trabajo comenzará con una extracción masiva de *tweets*, los cuales fueron categorizados por mes para luego hacer una limpieza de los datos, procurando mantener un *dataset* lo más fiel posible al problema abordado. Una vez obtenidos los datos refinados, se procedió a realizar análisis de sentimiento sobre los datos, con el fin de determinar percepciones positivas, neutras o negativas sobre la convención constituyente, así como también la presencia del discurso de odio y las emociones que puedan encontrarse vertidas en las opiniones entregadas por los votantes, como la felicidad o el enojo. Finalmente, a partir de los resultados se construyeron múltiples visualizaciones de datos que nos permiten entender de mejor manera la evolución del debate constitucional en Chile durante el 2022, sus actores principales y la forma en la que éstos interactuaron e influyeron en su entorno.

3.2. Selección de los datos

El *web scraping*, como ya fue explicado en el capítulo anterior, es el proceso de recolectar datos contenidos en páginas web mediante técnicas automatizadas, en este caso Twitter, una de las redes sociales más populares en el mundo que se caracteriza por centrarse más en texto que archivos multimedia como fotos o videos, a diferencia de Facebook, Instagram o TikTok, lo cual la convierte en la red social idónea de la cual poder extraer opiniones más fieles respecto del sentimiento de las personas.

En años anteriores, Twitter tenía disponible al público una API de libre acceso mediante la cual era posible extraer *tweets* para diferentes propósitos, como por ejemplo, con fines investigativos. Sin embargo, tras la llegada de Elon Musk, estas facilidades comenzaron lentamente a cambiar, modificando la API existente y presentando su segunda versión, la cual sólo es accesible mediante pago y con un límite mensual de *tweets* extraíbles muy baja para los propósitos de este documento.

De esta forma, fue necesario encontrar nuevos métodos basados en *web scraping* que permitieran extraer la información base sobre la cual trata este trabajo. Una forma de lograr esto, es, mediante bibliotecas de código abierto como BeautifulSoup y Selenium simular la navegación de una persona natural sobre la plataforma de Twitter, la cual puede filtrar *tweets* en base a sus necesidades y finalmente extraerlos para su uso. Sin embargo, nuevas modificaciones se fueron desarrollando con el pasar del 2023, con las cuales se limitó la cantidad máxima de *tweets* que puede ver un usuario diariamente, haciendo de esta metodología extremadamente tediosa y extensa, por ende, inviable.

Teniendo todo esto en cuenta, finalmente la propuesta de la solución se construyó a partir de Snsrape, un scraper construido en Python, especializado para *social networking services* (SNS), entre los cuales se encuentran las redes sociales más populares y usadas a nivel mundial, como Facebook, Instagram, Reddit y Twitter. Snsrape, al igual que la metodología propuesta anteriormente basada en BeautifulSoup y Selenium, "navega" por la plataforma de Twitter extrayendo los *tweets* deseados, con la diferencia fundamental de que Snsrape maneja múltiples sesiones anónimas dentro de la página web que le permiten navegar libremente por dicha red social sin limitaciones de ningún tipo, pudiendo consultar por *tweets* antiguos y actuales, filtrar por palabras específicos, en un rango de fechas determinado, e incluso la capacidad de incluir o excluir aquellos *tweets* que sean respuestas de otros, funcionalidades que fueron de utilidad en el proceso.

Para asegurar mejores resultados, la estrategia se basó en prescindir de la mayor cantidad de características posibles, siempre y cuando éstas fuesen consideradas como no relevantes para el estudio, lo cual se determinó en dos instancias, antes y después de la extracción, ya que gracias a la documentación de Snsrape, se sabe que algunas de las características que vienen incluidas con cada *tweet* son las siguientes:

- Url: Corresponde a la dirección url del *tweet*.

- **Date:** Corresponde a la fecha en la que fue publicada el *tweet*.
- **RawContent:** Corresponde al cuerpo del *tweet* sin formato, es decir, texto plano.
- **RenderedContent:** Corresponde al cuerpo renderizado del *tweet*, es decir, lo que finalmente se ve en Twitter una vez fue publicado.
- **Id:** Corresponde al ID interno del *tweet* registrado en Twitter.
- **User:** Corresponde al *username* único de la persona que publicó el *tweet*.
- **ReplyCount:** Corresponde a la cantidad de respuestas que fueron registradas en el *tweet* al momento de su extracción.
- **LikeCount:** Corresponde a la cantidad de *likes* o me gusta que fueron registrados en el *tweet* al momento de su extracción.
- **QuoteCount:** Corresponde a la cantidad de *quotes* o menciones que fueron registradas en el *tweet* al momento de su extracción.
- **RetweetCount:** Corresponde a la cantidad de *retweets* que fueron registrados en el *tweet* al momento de su extracción.
- **Lang:** Corresponde al lenguaje en el cual está escrito el *tweet*.
- **Links:** Corresponde a la lista de links presentes en el *tweet*.
- **Media:** Corresponde al contenido multimedia presente en el *tweet*.
- **Coordinates:** Corresponde a las coordenadas geográficas desde donde fue publicado el *tweet*.
- **Hashtags:** Corresponde a la lista de *hashtags* presentes en el *tweet*.
- **Vibe:** Corresponde a la "vibra" del *tweet*.
- **Pinned:** Característica que indica si el *tweet* extraído se encuentra fijado o no.

Antes de la extracción se planteó la posibilidad que atributos como Media no aportaran mayor información y finalmente fueron descartados dado que su análisis no se condice con el enfoque de esta solución, y correspondería a un estudio totalmente aparte con metodologías diferentes y requiriendo de un proceso mucho más extenso que el descrito en el presente documento, dado que el foco esta investigación se basó principalmente en el contenido en texto del *tweet* y el sentimiento contenido en éste y no el contenido multimedia que pueda venir asociado. De la misma manera, características como Pinned no aportaron mayor información y son potencialmente prescindibles dado que es difícil y ambiguo determinar el impacto que podría tener que un *tweet* se encuentre fijado o no, sabiendo que esto es

decisión del usuario que lo publicó y puede deberse a diversos motivos que no son obtenibles mediante ninguna herramienta o metodología disponible en la actualidad. Por ende, a partir de la lista anterior, se definieron los siguientes atributos para considerar en la primera extracción:

- *Date* (fecha)
- *User* (usuario)
- *RawContent* (el cuerpo del *tweet*)
- *RenderedContent* (el cuerpo del *tweet* renderizado)
- *ReplyCount* (cantidad de respuestas)
- *RetweetCount* (cantidad de *retweets*)
- *LikeCount* (cantidad de *likes*)
- *QuoteCount* (cantidad de menciones)
- *Hashtags*
- *Vibe*
- *Coordinates*

Con la definición de atributos ya establecida, lo siguiente fue determinar el filtrado de los *tweets* recolectados. Así, el filtro preliminar en Snsrape fue definido con las siguientes reglas:

- Todos los *tweets* extraídos deben estar en español. Esto debido a que el algoritmo que utilizado (PySentimiento) está diseñado para este lenguaje. Además, como el contexto de esta investigación trata del debate constitucional en Chile, es natural que la mayoría de las opiniones encontradas se encuentren en el idioma oficial de nuestro país, es decir, el español.
- Los *tweets* extraídos deben contener las palabras 'convención' o 'constituyente' o ambas juntas, es decir, no son excluyentes.
- La extracción se realizó mes por mes y fue almacenada en archivos independientes, es decir, todos los *tweets* correspondientes a un determinado mes se encontraron guardados en un sólo archivo y corresponden al lapso comprendido entre el primer día de dicho mes y el último. Se definió esta regla en base a que se realizó un análisis de la evolución del debate constitucional, que de haberse hecho por día o por semana habría resultado demasiado largo y fuera del alcance de este trabajo.

- Se excluyeron aquellos *tweets* que contuvieran links en su interior, debido a que la presencia de éstos en el análisis de sentimiento pudo haber afectado negativamente al resultado y habría hecho necesario el desarrollo de un algoritmo adicional que identificara *tweets* con enlaces en su interior para luego eliminarlos y dejar sólo el cuerpo con texto, lo cual, si bien era factible de realizar, inducía a una pérdida de contexto si es que el *link* presente, por ejemplo, hacía referencia a una noticia.

De esta manera, fue posible construir un dataset compuesto por aproximadamente 400.000 *tweets* comprendiendo el lapso entre enero y septiembre del año de 2022, el cual en desglose, contiene alrededor de 50.000 *tweets* por mes exceptuando septiembre debido a que el plebiscito constitucional tomó lugar el día 4 de ese mes, por lo cual se consideró sólo hasta ese día para la recolección de datos, es decir, se obtuvo la cifra mensual considerando desde enero hasta agosto incluyendo los primeros 4 días septiembre como adición.

En el transcurso de la ejecución de la solución se planteó la interrogante de qué tanto afectaría a la calidad del dataset el hecho de que este contenga *tweets* originales y sus respuestas, entendiendo que la presencia de éstas últimas podría entregar una falsa sensación de una percepción positiva, neutral o negativa dependiendo del volumen encontrado, es decir, aquellos *tweets* con mayor cantidad de respuestas podrían desequilibrar los resultados obtenidos en el análisis posterior.

Es por esto que se decidió construir un segundo dataset que se compone únicamente de *tweets* originales sin sus respuestas, bajo la hipótesis de que un conjunto de datos con estas características podría ser más representativo del sentimiento general del votante chileno encontrado en Twitter. Además, la disponibilidad de ambos *datasets* hace posible que, al efectuar el análisis de sentimiento y posteriormente obtener sus resultados, se pueda realizar análisis comparativo que finalmente permita determinar si es que la hipótesis inicial bajo la cual se decidió construir el *dataset* sin respuestas era cierta o no.

Gracias al sistema de filtrado de Snsrape, el cual también permite excluir de la extracción aquellos *tweets* que no sean originales, es decir, que sean en respuesta a otro ya existente, se logró construir el segundo *dataset* manteniendo las reglas que fueron utilizadas para la construcción del primero. Dado que la cantidad de *tweets* originales es menor que la cantidad de la suma de los mismos más sus respuestas, su volumen fue considerablemente menor, alcanzando aproximadamente una cifra de 100.000 *tweets*. Sin embargo, el filtrado no fue tan preciso y trajo consigo muchos *tweets* que no eran útiles para el estudio y por ende pudieron afectar negativamente a los resultados y conclusiones obtenidas, para lo cual fue necesario implementar un algoritmo de limpieza capaz de detectar aquellos *tweets* que no se condicen con el contexto en cual se llevó a cabo esta investigación.

3.3. Preprocesamiento de datos

Una vez obtenidos los datos, siguiendo la metodología KDD, fue necesario realizar un análisis previo abarcando dos focos principales de la naturaleza del *dataset* construido: la estructura de los datos y el contenido.

Como se vio en el marco teórico, un *tweet* se compone de más que sólo el texto que lleva en su interior, sino que también tiene diversas características cualitativas y cuantitativas asociadas a éste que resultaron ser útiles para esta propuesta de solución, mientras que algunas fueron obviadas y excluidas del análisis debido a que no fueron relevantes para el contexto y no aportaban a un mejor resultado.

Dentro de dichas características, es posible encontrar algunas de carácter cualitativo como el contenido mismo del *tweet*, es decir, el cuerpo, así como también lo que quiere emitir el usuario que lo publicó, y también características cuantitativas como la cantidad de *likes* o *retweets* que se van registrando en Twitter una vez que ya fue publicado el *tweet* y que se actualiza cada vez que recibe una reacción o respuesta nueva.

Si bien Snsrape es capaz de traer todas estas características en la extracción y muchas más, el análisis fue importante para determinar si realmente todas eran necesarias y útiles para el estudio, ya que muchas de éstas no aportaban información debido a diferentes motivos, ya que al momento de la extracción estos datos no estaban disponibles, y algunas características se encontraban obsoletas y ya no contenían información dentro de éstas.

Luego de obtener los datos mediante Snsrape, se procedió a realizar el análisis de la estructura de éstos con el fin de decidir la información que era relevante estudiar para el contexto de esta investigación y fuese capaz de aportar mayor profundidad a ésta.

Los campos obtenidos en la extracción de cada *tweet* fueron los siguientes:

- *Date*: Corresponde a la fecha en la que fue publicado el *tweet* en el formato yyyy-MM-dd hh:mm:ss.
- *User*: Corresponde al nombre único de la cuenta del usuario que publicó el *tweet*.
- *RenderedContent*: Corresponde al cuerpo renderizado del *tweet*.
- *RawContent*: Corresponde al contenido en texto del *tweet* publicado.
- *ReplyCount*: Corresponde al recuento de respuestas obtenidas por el *tweet* al momento de su extracción.
- *RetweetCount*: Corresponde al recuento de *retweets* obtenidos por el *tweet* al momento de su extracción.
- *LikeCount*: Corresponde al recuento de *likes* o 'me gusta' obtenidos por el *tweet* al momento de su extracción.

- *QuoteCount*: Corresponde al recuento de *quotes* o menciones obtenidas por el *tweet* al momento de su extracción.
- *Hashtags*: Corresponde a una lista de los *hashtags* presentes en el *tweet* extraído.
- *Vibe*: Corresponde a la "vibra" presente en el texto del *tweet* extraído.
- *Coordinates*: Corresponde a las coordenadas aproximadas desde donde fue publicado el *tweet*.

De los cuales *Date*, *Vibe*, *RenderedContent* y *Coordinates* fueron desestimados del análisis. En el caso de la fecha (*Date*), la razón de su exclusión radica en que la metodología empleada para la extracción de *tweets* fue efectuada de manera mensual y se encuentra etiquetada debidamente, por lo cual este campo no aporta información adicional; además, el estudio de la evolución del debate constitucional fue realizado mes por mes, debido a que el hacerlo de manera diaria extendería demasiado el alcance de esta investigación. Por su parte, *Vibe* no fue considerado debido a que ninguna de las extracciones trajo algún valor alguno, por lo cual no aportaba ninguna información a lo que se deseaba realizar.

En el caso de *RenderedContent*, su equivalente *RawContent* mostró una mayor facilidad para ser trabajado en las fases siguientes del proceso dado que corresponde a texto plano que es fácilmente trabajable y modificable, por ende, al no aportar mayor información útil para los propósitos de este estudio, finalmente fue descartado. *Coordinates* al igual que *Vibe* no arrojó información alguna, razón por la cual quedó automáticamente excluido. La no presencia de este campo dentro del estudio imposibilita la opción de efectuar un análisis geográfico que podría aportar información valiosa a los resultados de esta investigación.

3.4. Transformación de datos

El segundo aspecto de la naturaleza de los datos obtenidos corresponde a su contenido mismo, en el sentido de saber si es que el contenido en texto de cada *tweet* era relevante para el estudio o no. Como se dijo en el punto anterior, referente a la extracción de los datos mediante Snsrape, muchos de los *tweets* obtenidos podían no corresponder al contexto del debate constitucional en Chile durante el 2022 y por ende afectar negativamente al resultado que esta propuesta entregó.

La tarea de limpiar el *dataset* debió tomarse con cautela para eliminar en su mayoría aquellos datos que no eran útiles para la solución, procurando también minimizar la pérdida de aquella información que sí era relevante para el análisis. La revisión manual de los datos fue fundamental para identificar el patrón comprendido de los *tweets* que debieron ser excluidos del *dataset*, por lo que finalmente, el algoritmo de limpieza del *dataset* se enfocó en tres fuentes de información que no guardaban relación directa con el debate constituyente en Chile durante el 2022:

- Llamado a asamblea constituyente en Perú (Abril 2022): Propuesta por el ex presidente del Perú, Pedro Castillo, impulsó una gran cantidad de opiniones y polémicas en Twitter debido a la álgida crisis política que se vivía en dicho país, que terminó con la destitución de Castillo.
- Protestas en Colombia (2021-2022): Como respuesta a las reformas tributarias y de salud impulsadas por el aquel presidente Iván Duque, las protestas masivas se tomaron las calles de Colombia, siendo el ciberactivismo una vía alternativa de protesta que se tomó una buena parte de Twitter durante los años 2021 y 2022, mencionando la posibilidad de una reforma constitucional de la cual cuyo actor clave fue Gustavo Petro, actual presidente de Colombia.
- *Tweets* genéricos que hablan de una 'convención': Como el filtrado y selección de *tweets* se hizo en base a palabras claves, una cantidad menor de éstos hablan acerca de convenciones como un sustantivo común o sobre acuerdos internacionales históricos como la Convención de Ginebra.

Teniendo en cuenta estas tres fuentes, se comenzó en la búsqueda de palabras claves que permitieran encontrar aquellos *tweets* que se refirieran a alguno de estos tópicos. Para los primeros dos casos, se definieron como palabras claves los nombres y apellidos de los actores claves de dichos conflictos, como el caso de Pedro Castillo y Wladimir Cerrón en Perú y Gustavo Petro en Colombia, así como también algunos *hashtags* utilizados para referirse al momento en específico que hicieron más fácil la identificación de dichos *tweets*. Así mismo, la mención a medios de comunicación locales, fueron también una pieza clave de la cuál se alimentó el algoritmo de limpieza para eliminar dichos *tweets* del *dataset*. Para el tercer y último caso, los nombres de acuerdos históricos como la Convención de Ginebra u eventos masivos como la Comic Con fueron utilizados como *keywords* para intentar minimizar la presencia de este tipo de *tweets* en la fuente final de datos.

Con las palabras claves ya definidas, el algoritmo se encargó de eliminar la mayoría de los *tweets* que no correspondían a la Convención Constituyente, reduciendo la cantidad inicial de aproximadamente 400.000 *tweets* a unos 320.000, es decir un 20 por ciento menos del *dataset* original. En el caso del segundo *dataset*, el cual no tenía respuestas, la disminución fue menor, quedando con una cifra total de aproximadamente 90.000 *tweets*.

3.5. Análisis de Sentimiento, Emoción y Detección de Discurso de Odio (PySentimiento)

Tal como se habló en el marco teórico, el Análisis de Sentimiento es una rama del Procesamiento del Lenguaje Natural, que abarca el estudio computacional, emociones y opiniones expresadas en texto. Como los *tweets* se componen mayoritariamente de texto que contiene

una opinión vertida por el usuario que lo publicó, el análisis de sentimiento fue una herramienta bastante potente y útil para dilucidar el parecer de la población en general respecto a la Convención Constituyente.

El análisis de sentimiento puede implementarse de muchas maneras dependiendo de la naturaleza de los datos, el lenguaje utilizado y los recursos disponibles. En la actualidad, el análisis de sentimiento basado en *machine learning* es el más preciso de todos los métodos y el más utilizado, que gracias a la inteligencia artificial y la cantidad creciente de algoritmos y modelos de libre uso, es capaz de interpretar el texto presente y entregar un porcentaje de predicción del sentimiento presente en la oración entregada.

Dentro de los algoritmos de análisis de sentimiento basados en *machine learning* disponibles, existen los de tipo supervisados y no supervisados, cuya diferencia radica principalmente en la necesidad de un etiquetado previo de los datos previo a su análisis: el supervisado requiere de una clasificación manual de los *tweets*, mientras que el no supervisado cuenta con un modelo pre entrenado, el cual permite predecir con precisión el sentimiento en una oración sin la necesidad del etiquetado de los datos.

Dentro de los algoritmos no supervisados disponibles, fue importante tomar en cuenta el lenguaje en el cual venía escrito el *dataset*. Dado que se habla del debate constitucional en Chile, el idioma predominante es el español, lo cual hizo necesaria la búsqueda de un algoritmo que estuviera previamente entrenado en ese lenguaje. Como la mayoría de los modelos disponibles han sido desarrollados para el inglés, una opción podría haber sido traducir los *tweets* al inglés para luego procesarlos, aunque el riesgo de pérdida clave de información era inminente, sobre todo teniendo en cuenta la cantidad de modismos con las que cuenta el español y el lenguaje informal que muchas veces se emplea en redes sociales como Twitter.

Así, el candidato seleccionado para este estudio fue PySentimiento, un algoritmo no supervisado de análisis de sentimiento preentrenado para el español, publicado por primera vez el año 2021 por Juan Manuel Pérez, Juan Carlos Giudici y Franco Luque como un "*toolkit* multilinguaje desarrollado en Python para tareas de análisis de sentimiento y SocialNLP", el cual fue entrenado con modelos preexistentes como BERTweet, un modelo preentrenado para *tweets* en inglés y que se convertiría en la base de modelos pre entrenados posteriores en otros idiomas.

Un año más tarde, en 2022, dos de los autores originales de PySentimiento, Pérez y Luque, se unirían a Laura Alemany y Ariel Furman para desarrollar RoBERTuito, un modelo pre entrenado para lenguaje de redes sociales en español, con el cual finalmente se llevó a cabo esta investigación. En adición, otros dos modelos fueron utilizados con el fin de ampliar el alcance de este estudio y entregar más información acerca del debate constitucional en Chile durante el 2022: EmoEvent, modelo multilinguaje enfocado en la detección de emociones en texto, y SemEval, modelo especializado para la detección de discurso de odio en inglés y español.

Teniendo en consideración lo expuesto en el párrafo anterior, el análisis comprendido en esta

etapa del proceso se dividió en tres grandes partes:

- Análisis de sentimiento
- Análisis de emociones
- Detección de discurso de odio

3.5.1. Análisis de sentimiento

El volumen de los *datasets* obtenidos anteriormente abre la posibilidad de realizar múltiples análisis desde diferentes aristas, con el fin de obtener la mayor cantidad de información posible que luego pueda ser interpretada en los resultados que nos den una visión más amplia de lo que fue el debate constitucional del 2022. Teniendo esto en cuenta, se desarrolló una estrategia acorde a estos propósitos, con hallazgos relevantes y con una narrativa consistente que logre despertar el interés de las personas que lean este documento.

De esta manera, la estrategia que se adoptó en esta primera etapa fue abarcar desde lo general hasta lo más particular, con la finalidad de obtener un análisis profundo y multicapa, en el cual cada una de ellas se complementen y logren, en conjunto, formar un amplio panorama que pueda ser interpretado y estudiado a futuro.

El análisis más general que se puede efectuar entonces, es aplicar análisis de sentimiento directamente sobre los datos obtenidos, sin realizar ningún tipo de segmentación previa, más allá de la ya existente separación por meses debido a la metodología empleada en la extracción de los datos. Para realizar esto, primero se comenzó con el *dataset* de menor volumen, es decir, los *tweets* originales sin respuesta, dado que su tiempo de procesamiento fue menor y por ende se puede obtener de manera inmediata una primera visión de lo que fue el debate constitucional en Twitter.

El equipo utilizado para esta etapa del experimento consistió en un notebook equipado con las siguientes especificaciones:

- CPU: Intel Core i7-10870H de 8 núcleos y 16 hilos con una frecuencia base de 2.20 GHz, extendible hasta los 5 GHz
- Memoria RAM: 16 GB DDR4 @ 3200 MHz, dual-channel
- GPU: Nvidia RTX 2060 Mobile, 6 GB GDDR6
- Almacenamiento: 1 TB SSD M.2
- Sistema Operativo: Windows 10 Home

Dado que PySentimiento es un algoritmo basado en *machine learning* para realizar análisis de sentimiento, las especificaciones señaladas previamente son suficientes para llevar a cabo esta etapa a una velocidad aceptable. Sin embargo, debido a problemas de *drivers*, no fue posible realizar procesamiento con los núcleos CUDA presentes en la tarjeta gráfica descrita anteriormente, por lo cual se debió confiar netamente en la capacidad del procesador. Teniendo en cuenta que el procesamiento por núcleos CUDA es considerablemente más rápido, se redujo drásticamente la velocidad a la cual se llevó a cabo este proceso, aunque no fue impedimento para continuar con éste.

Así, se procedió a realizar análisis de sentimiento por cada *sub dataset* mensual, cuyo detalle se presenta en la Tabla 1.

Mes	Cantidad de tweets	Tiempo de procesamiento
Enero	14.869	13:29
Febrero	12.936	11:29
Marzo	9.751	08:48
Abril	12.139	11:12
Mayo	10.188	09:37
Junio	8.467	07:49
Julio	10.956	10:23
Agosto	8.210	07:48

Tabla 1: Volumen del *dataset* sin respuestas y su tiempo de procesamiento.
Fuente: Elaboración Propia.

Dado que los resultados obtenidos distan considerablemente de la tendencia mostrada en la Figura 1, fue necesario realizar un proceso adicional que validara la efectividad y precisión de PySentimiento y RoBERTuito para esta tarea, lo cual se hizo realizando una separación entre la capa de *testing* y la capa de entrenamiento, con el fin de medir la cantidad de verdaderos y falsos positivos y negativos. Con los resultados obtenidos, se determinó con exactitud la precisión del modelo utilizado.

Una vez hecha la separación, la matriz de confusión resultante se puede ver de acuerdo a lo presentado en la Figura 7.

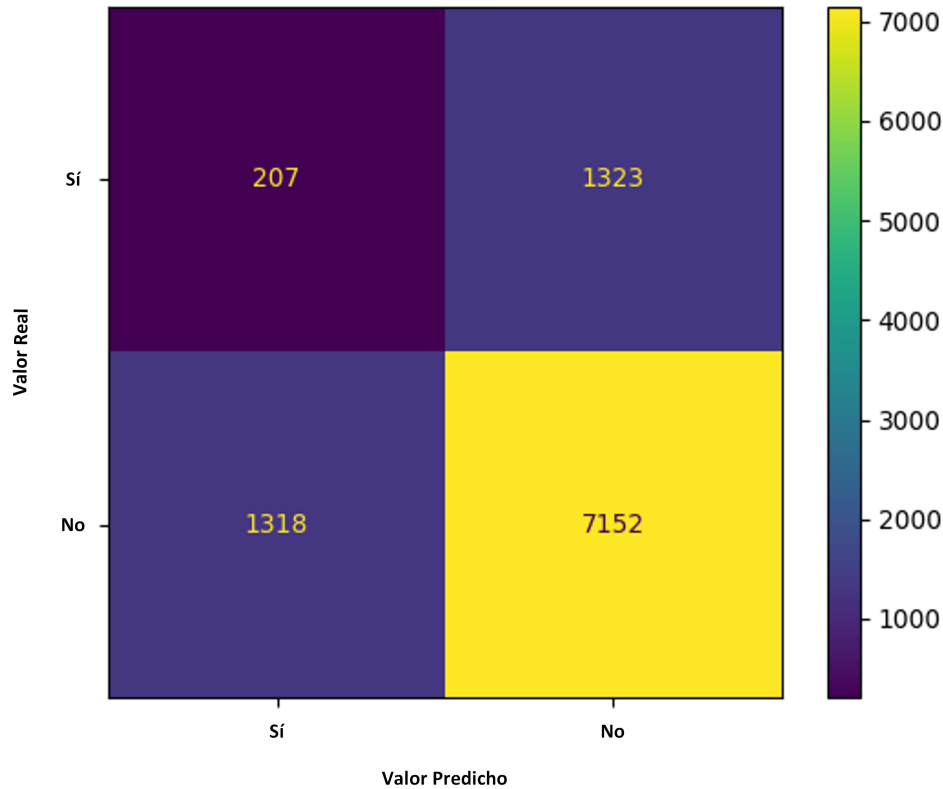


Figura 7: Matriz de confusión para análisis de sentimiento.
Fuente: Elaboración propia.

Como fue explicado en el marco teórico, la cantidad de VP y VN es directamente proporcional a la precisión del modelo, es decir, una mayor cantidad de predicciones correctas se traduce en una mejor calidad, lo cual lógicamente es acertado. Sin embargo, este método también toma en cuenta la detección de FP y FN , los cuales, a diferencia de los primeros, son inversamente proporcionales a la precisión, por ende, es importante que estas cifras sean las menores posibles. La matriz obtenida muestra un total de 7.152 VP y 207 VN los cuales se interpretan como niveles sobre lo aceptable para un modelo de estudio del lenguaje natural, mientras que presenta 1.323 FP y 1.318 FN , es decir, predicciones incorrectas, las cuales a su vez son números bajos para un modelo de esta naturaleza. Con todo lo expuesto anteriormente, es posible determinar una exactitud de 73,59% para el modelo y un 83,49% de precisión, cumpliendo un nivel de calidad del modelo suficiente para avanzar sin la necesidad de probar otros modelos. De la misma manera, este hallazgo refuta en parte la hipótesis inicial de que el apoyo a la convención constituyente fue más alto que su rechazo y luego la tendencia comenzó a invertirse, al menos en redes sociales. Con este punto de partida, y una nueva mirada del proceso constituyente, se continuó con los siguientes pasos.

Prosiguiendo con la metodología planteada, se repitió el proceso previamente señalado con el segundo *dataset*, es decir, el que contiene *tweets* originales y sus respuestas, principalmente para probar o refutar la hipótesis planteada en la propuesta de la solución acerca del impacto que puede tener en la percepción general de la convención constituyente la presencia de éstas. Cabe destacar que, si los resultados no variaban considerablemente en base al análisis anterior se procedería con la metodología sólo aplicada a este *dataset*; en caso contrario, se seguiría realizando el estudio con ambos *datasets* con el fin de dimensionar el impacto en la presunta diferencia entre ambos conjuntos de datos.

Utilizando el mismo equipo, cuyas especificaciones fueron detalladas con anterioridad, la cantidad de *tweets* analizados y el tiempo tomado se muestran en la Tabla 2.

Mes	Cantidad de tweets	Tiempo de procesamiento
Enero	35.346	30:11
Febrero	44.879	37:03
Marzo	39.902	35:36
Abril	32.160	28:03
Mayo	43.496	38:23
Junio	39.070	32:57
Julio	43.412	38:34
Agosto	42.623	36:27

Tabla 2: Volumen del *dataset* con respuestas y su tiempo de procesamiento.
Fuente: Elaboración Propia.

Al igual que para el procesamiento del primer *dataset*, el tiempo de procesamiento fue proporcional a la cantidad de *tweets* procesados, tardando mucho más tiempo en este análisis debido a la diferencia considerable en el volumen de datos. Con una velocidad promedio de 19,3 *tweets* por segundo, este análisis tomó un tiempo menor de análisis por *tweet* comparado con el primero, aunque tomó una cantidad mayor de tiempo efectivo debido a la cantidad de *tweets* analizados.

Luego de obtenidos la clasificación de los *tweets*, en la Tabla 3 se muestra la percepción de los usuarios de Twitter sobre la convención constituyente en un *dataset* con respuestas incluidas. Como se puede apreciar en dicha tabla, el análisis realizando bajo el *dataset* con respuestas entrega una percepción sutilmente diferente de la obtenida anteriormente, especialmente en las positivas y negativas. La percepción positiva presentó un máximo más alto en el mes de enero, con un 8,51 %, superando al 7,66 % marcado en julio en el análisis anterior, aunque también se registró un nuevo mínimo: 4,78 % en el mes de abril, 0,84 % menor que el mínimo encontrado en el primer *dataset*, que correspondía al mes de enero. De la misma manera, la percepción negativa de la convención constituyente alcanzó un nuevo máximo en febrero con un 72,44 %, 2,4 % mayor que la percepción negativa detectada en abril bajo el *dataset* anterior. Repitiendo la metodología aplicada anteriormente, se obtuvo la percepción promedio para cada categoría: *tweets* positivos: 6,14 %, *tweets* neu-

trales 26,34 % y tweets negativos: 67,52 %. En resumen, el porcentaje de tweets positivos disminuyó en un 0,61 %, los tweets detectados como neutrales bajaron 1 %, mientras que la percepción negativa fue la única que aumentó, haciéndolo con un 1,61 %.

Mes	Positivo(%)	Neutral (%)	Negativo (%)
Enero	8,51 %	31,37 %	60,12 %
Febrero	4,96 %	22,60 %	72,44 %
Marzo	5,22 %	27,21 %	67,57 %
Abril	4,78 %	23,18 %	72,04 %
Mayo	5,85 %	25,13 %	69,02 %
Junio	5,69 %	27,76 %	66,54 %
Julio	6,89 %	27,08 %	66,03 %
Agosto	7,24 %	26,39 %	66,37 %

Tabla 3: Resultado del análisis de sentimiento aplicado al *dataset* con respuestas
Fuente: Elaboración Propia.

Los resultados obtenidos en este nuevo análisis dan cuenta de que la percepción predominante, la negativa, fue la única que aumentó sus números, lo cual es consistente con la hipótesis planteada anteriormente sobre el peso que puede arrastrar un *tweet* de determinada posición y cómo esto finalmente puede afectar a la percepción general. Sin embargo, la diferencia no es significativa, y si bien los resultados comprueban la hipótesis propuesta, no hacen de éste un *dataset* menos representativo; al contrario, constituye una mirada más fiel a la realidad del debate constitucional, dado que se toma en consideración que el interés de las personas por opiniones negativas, también representa un tipo de percepción negativa más pasiva pero no descartable. Es por esto que finalmente, los análisis que se presenten de aquí en adelante fueron confeccionados a partir del *dataset* que incluye *tweets* originales y sus respectivas respuestas.

En párrafos anteriores se mencionaron métricas conocidas como la positividad, neutralidad y negatividad, las cuales se calculan en base al peso asignado por PySentimiento a cada percepción al momento de su clasificación, es decir, un *tweet* calificado como negativo en realidad no es completamente negativo, lo cual también aplica en el caso positivo o neutral. Por ejemplo, un *tweet* puede obtener los siguientes pesos antes de ser clasificados:

- Positivo: 10,13 %
- Neutral: 36,81 %
- Negativo: 53,06 %

Bajo lo cual sería clasificado como negativo, al ser el peso más alto de los tres. Sin embargo, también se puede concluir que dicho *tweet* es 46,94 % no negativo, información que se

pierde al solamente analizar las clasificaciones finales entregadas por PySentimiento. Como dichos pesos también son entregados por el modelo al momento de su ejecución, fue posible construir la Tabla 3.

Mes	Positividad(%)	Neutralidad (%)	Negatividad (%)
Enero	11,25 %	31,81 %	56,94 %
Febrero	7,51 %	25,85 %	66,64 %
Marzo	7,96 %	29,18 %	62,86 %
Abril	7,33 %	26,18 %	66,49 %
Mayo	8,48 %	27,63 %	63,89 %
Junio	8,57 %	29,72 %	61,70 %
Julio	9,71 %	29,23 %	61,05 %
Agosto	9,93 %	28,50 %	61,58 %

Tabla 4: Positividad, neutralidad y negatividad presente en el *dataset* con respuestas
Fuente: Elaboración Propia.

Si bien las tendencias no se ven alteradas, es decir, la percepción negativa sigue siendo la más predominante y la positiva la menor, los porcentajes se distribuyen de una manera más uniforme, viéndose beneficiada la positividad, que en promedio alcanza un 8,84 %, es decir, un 2,7 puntos mayor que el porcentaje de *tweets* calificados como positivos obtenido en el análisis anterior, mientras que la negatividad disminuye de una forma más significativa, con 4,88 puntos menos respecto al 67,52 % anterior. La neutralidad también gana un leve aumento respecto del 26,34 % original, llegando a 28,51 %, o sea, una diferencia de 2,17 puntos porcentuales. Cabe destacar también, que bajo este análisis el mes de enero presenta una mayor consistencia con la hipótesis de que el apoyo a la convención constituyente fue disminuyendo gradualmente desde comienzos del 2022, dado que es el mes con mayor positividad (11,25 %) y la menor negatividad (56,94 %), aunque igualmente dista de lo que se esperaba en un comienzo.

Con esto, el análisis más general de la percepción de los usuarios de Twitter sobre la convención constituyente queda concluido, entendiendo que las opiniones negativas predominaron dentro de los *datasets* analizados y que aquellas que se manifestaban a favor de ésta fueron las menores, en un grado bastante menor al esperado inicialmente, previo a la obtención de los primeros resultados. Siguiendo el enfoque de lo general a lo particular, es momento de ahondar un poco más en los aspectos que explican las razones detrás de esta primera percepción general.

En el marco teórico se habla brevemente acerca de la estructura de un *tweet*, componente principal de los *datasets* estudiados, dentro de la cual destaca la presencia de los *hashtags*, palabras claves propias de un *tweet* cuya definición y terminología nació en la red social Twitter. Con múltiples propósitos, los *hashtags* a menudo pueden dar un resumen de lo que se quiere decir en un *tweet* y precisar una idea, lo cual a su vez muchas veces puede ser utilizado como un método de clasificación, los llamados *trending topics* que pueden verse a diario

al utilizar Twitter desde su plataforma, indicándole al usuario cuáles son los temas de conversación más recurrentes en el mundo, en su país o de acuerdo a sus intereses personales. Es necesario precisar que los *hashtags* son añadidos voluntariamente por las personas, por ende, pueden no existir en un *tweet*, lo cual no impide realizar un análisis adicional en base a los cuales sí contengan. En esta etapa del proceso se enumeraron aquellos *hashtags* más recurrentes dentro del *dataset*, que luego se clasificaron de acuerdo al sentimiento encontrado en su *tweet* de procedencia.

Dentro de los atributos obtenidos en la extracción hecha mediante Snsrape, se obtiene una lista de los *hashtags* utilizados por cada *tweet*, en caso de que éste los contenga. Contabilizando cada una de estas listas por cada *tweet* y por cada mes, se tiene que los *hashtags* más recurrentes en el *dataset* analizado, son los que se encuentran en la Tabla 5.

<i>Hashtag</i>	Apariciones
#Rechazo	1.815
#ConvencionConstitucional	1.436
#RechazoTransversal	1.082
#RechazoDeSalida2022	1.035
#Apruebo	1.026
#RechazoDeSalida	895
#ConvencionCulia	711
#Apruebo4deSeptiembre	526
#rechazo	496
#RechazoElMamarrachoComunista	493

Tabla 5: 10 *Hashtags* más recurrentes encontrados en el *dataset*.
Fuente: Elaboración Propia.

Esta tabla da cuenta de que 6 de los 10 *hashtags* más recurrentes incluyen la palabra “Rechazo”, que es el nombre de la opción ganadora del plebiscito constitucional del 2022, cuyos votantes se manifestaron en contra de la propuesta de constitución presentada por la convención constituyente, es decir, en contra de su trabajo en general y que cuya opinión en Twitter podría calificarse como “negativa”. Sólo 2 de los 10 *hashtags* presentados contienen la palabra “Apruebo”, la opción perdedora, cuya cantidad de apariciones es ampliamente superada por el resto de los *hashtags* relativos al Rechazo, 1.552 apariciones conjuntas en contra de las 7.252 de la contraparte. El *hashtag* #ConvencionCulia hace uso de una grosería propia del español chileno que usualmente indica algo malo cuando se usa coloquialmente, por lo cual también podría sumarse a la lista de *hashtags* relativos al Rechazo, ampliando aún más la brecha con el Apruebo. El *tweet* restante, #ConvencionConstituyente, es consistente con la naturaleza de los datos estudiados y no es posible atribuírselo a un bando en específico, dado que puede tratarse de *tweets* de carácter informativo o con opiniones más neutrales sobre dicho organismo y el proceso.

Para comprobar si efectivamente los *hashtags* relativos al Rechazo están contenidos en

tweets que presentan opiniones negativas de la convención constituyente, lo mismo para el caso del Apruebo y la naturaleza del *hashtag* #ConvencionConstituyente, fue necesario ir a la clasificación del sentimiento que cada *tweet* obtuvo previamente en el proceso. Como cada *tweet* quedó asociado a un sentimiento, fue posible confeccionar la Tabla 6.

Hashtag	Sentimiento	Apariciones
#Rechazo	Negativo	1.470
#Apruebo	Positivo	974
#RechazoTransversal	Negativo	943
#RechazoDeSalida2022	Negativo	897
#ConvencionConstitucional	Negativo	849
#RechazoDeSalida	Negativo	777
#ConvencionCulia	Negativo	606
#Apruebo4deSeptiembre	Positivo	506
#AprueboDeSalida	Positivo	478
#RechazoElMamarrachoComunista	Negativo	461

Tabla 6: 10 *Hashtags* más recurrentes encontrados en el *dataset*, clasificados por sentimiento.

Fuente: Elaboración propia.

Ahora, con clasificación por sentimiento, es posible comprobar que todos los *hashtags* relativos al Rechazo efectivamente provienen de *tweets* clasificados con una percepción “negativa” de la convención constitucional, sumando un total de 4.548 apariciones, a las cuales se le suman las apariciones de los *hashtags* #ConvencionCulia, que también se comprueba su negatividad hacia la convención, dado que 606 de sus 711 apariciones fueron clasificadas como negativas, y de #ConvencionConstituyente, que en un inicio se pensó que podía ser un *hashtag* más neutral. Sin embargo, al aparecer 849 en *tweets* “negativos” pasa a clasificarse como un *hashtag* negativo en su mayoría, lo cual también es consistente con la tendencia identificada hasta ahora: la percepción de los usuarios de Twitter sobre la convención constitucional fue en su mayoría negativa. Los dos *hashtags* asociados al Apruebo fueron categorizados como positivos, coherente con lo esperado, los cuales sumaron un *hashtag* más en la forma de #AprueboDeSalida, añadiendo 478 apariciones a los *hashtags* “positivos”. Es importante mencionar que en los datos obtenidos no existe ningún *hashtag* asociado a un sentimiento neutral, sin embargo, esto se debe a que no se encuentra en el *top 10*, por lo cual no se muestra en esta sección pero sí más adelante cuando se presente la visualización de datos confeccionada en base a esta información.

Tomando como referencia los datos obtenidos en la Tabla 6, la distribución de los *hashtags* se presenta de la siguiente manera:

- Número de apariciones de *hashtags* “negativos”: 6.003
- Número de apariciones de *hashtags* “positivos”: 1.958

Que en porcentaje se traducen a un 75,4 % de *hashtags* clasificados como negativos y un 24,6 % como positivos. Hay que recalcar que dichas métricas fueron calculadas en base a la tabla 6 y no representan al universo total de *tweets* estudiados, por lo que los resultados podrían variar considerablemente al ampliar la muestra.

El estudio de los *hashtags* más recurrentes y los sentimientos asociados a sus *tweets* de procedencia ha ampliado un poco más la visión acerca de la percepción de las personas sobre la convención constitucional, confirmando que los *hashtags* relativos al Rechazo efectivamente vienen de opiniones negativas sobre el organismo redactor, mientras que los asociados al Apruebo se manifiestan a favor de éste. El hallazgo hecho sobre el *hashtag* #Convencion-Constitucional, donde la mayoría de apariciones fue desde *tweets* clasificados como negativos, comprueba la hipótesis de que el debate constitucional en sí, tuvo más opiniones negativas que positivas en su conjunto. Ahora bien, hay aspectos que aún deben analizarse, como por ejemplo, las personas que estuvieron detrás de los *tweets* y los *hashtags* publicados, para entender desde una posición más política y humana el objeto de estudio de este documento.

Recordando lo mencionado acerca de Snsrape, la herramienta utilizada para extraer los *tweets*, parte de los atributos seleccionados guardan relación con los usuarios que participaron, tanto aquellos que publicaron *tweets* como los que fueron mencionados. Es muy importante hacer esa diferencia, puesto que cualquier persona puede escribir cientos de *tweets* sobre un tema y ser el *twittero* número uno sobre el tópico. Sin embargo, pocas personas pueden decir lo mismo en menciones, es decir, ser los más mencionados cuando se habla de un tema, lo cual puede interpretarse como que su presencia marca influencia sobre los demás. Uno de los objetivos de este estudio es justamente determinar el grado de influencia de determinados usuarios, con el fin de identificar a aquellos que lograron influenciar en mayor medida el pensamiento de los demás. Esta última sección enfocada en el análisis de sentimiento tratará sobre el rol de las personas en el debate constitucional y se identificarán a aquellos personajes claves que fijaron el rumbo del mismo.

La primera parte de este análisis se enfocó en identificar a los usuarios que más *tweets* publicaron sobre la convención constituyente en el lapso comprendido entre enero y el plebiscito del 4 de septiembre del 2022, para luego hacer un estudio acerca de sus perfiles, siempre y cuando sean de carácter público, con el fin de determinar la postura que dichas personas sostuvieron durante el debate constitucional, pudiendo construir así un perfil promedio de los *twitteros* más activos dentro del proceso.

Analizando el *dataset*, los usuarios que más publicaron sobre la convención en el 2022 previo al plebiscito se muestran en la Tabla 7.

Usuario	Cantidad de tweets
@themasterxtc	1.141
@Pantera_Chileno	647
@silbersam	507
@Chagon16	472
@juanriv42678873	364
@Votamos__Todos	292
@AlejandraLisbo6	274

Tabla 7: 7 usuarios que más tweets publicaron sobre la convención constituyente entre enero y septiembre del 2022.

Fuente: Elaboración propia.

Ninguno de los usuarios presentados en la tabla anterior parece ser algún personaje relevante en el debate constitucional, así como un constituyente, un personaje político o un medio de comunicación, lo cual es consistente con lo planteado acerca de la diferencia entre los usuarios que más publican y los más mencionados. Sin embargo, para probar dicho planteamiento, fue necesario investigar a quienes se encontraban detrás de estos perfiles.

- @themasterxtc: Según su cuenta de Twitter, su nombre real es Sebastián Gosselin y se define a sí mismo como anticomunista y anti “zurdolandia”, haciendo referencia a la izquierda como bando político. Suele publicar en contra de la izquierda chilena y del gobierno actual de Gabriel Boric, mientras que se muestra a favor de personajes como el dictador Augusto Pinochet.
- @Pantera_Chileno: Cuenta eliminada u oculta en Twitter.
- @silbersam. En Twitter no aparece algún nombre asociado, pero es un usuario activo que suele *twittear* sobre política chilena. Se declara en contra del golpe militar así como también en contra del estallido social.
- @Chagon16: Apodado “Chagón”, es una persona que suele publicar en contra de la izquierda chilena, con una postura “anticomunista”, a los cuales ha llegado a llamar “hijos de puta”.
- @juanriv42678873: Llamado Juan Rivas, es un usuario que publica constantemente en contra del gobierno de Boric, tildándolo de “terrorista” y “ladrón”. En ocasiones, ha dicho que el gobierno del presidente Salvador Allende fue el peor de toda la historia de Chile.
- @Votamos__Todos: Son un medio independiente llamado Votamos Tod@s Noticias, que se dedicó principalmente a seguir la convención constituyente. Actualmente ya no se encuentra activo en Twitter.

- @AlejandraLisbo6: Llamada Alejandra Lisboa, es una usuaria de Twitter que usualmente *twittea* sobre política chilena, con una postura en contra de la izquierda, tildando al presidente Boric de “comunista” y “ladrón”.

El análisis de los perfiles de las personas que más publicaron sobre la convención constituyente durante enero y septiembre del 2022 entrega información sobre 6 de los 7 usuarios, puesto que @Pantera_Chileno ya no existe en Twitter o es un perfil al cual no es posible acceder. De los restantes, 4 tienen una marcada tendencia hacia la derecha o ultraderecha política, bandos que apoyaron al Rechazo e hicieron campaña activa en contra de la propuesta constitucional redactada por la convención, lo cual no es de extrañar considerando la alta percepción negativa sobre este organismo que fue detectada en este análisis. En cuanto a @silbersam es difícil determinar su postura sobre la convención en base a su perfil y últimas publicaciones puesto que no tiene una tendencia marcada en lo que *twittea*. Finalmente, @Votamos__Todos es un perfil inactivo de un medio independiente de noticias que se enfocó específicamente en el tópico principal de este estudio, del cual el análisis no logró identificar algún sesgo, por lo que se presume que sus *tweets* fueron meramente informativos, imparciales y objetivos.

Se repitió la metodología aplicada a los *hashtags* con el fin de comprobar o refutar las hipótesis respecto a las opiniones de los usuarios en estudio, es decir, se hizo uso de la clasificación previamente obtenida de los *tweets*, esta vez, obtenidos desde el usuario que los publicó.

Usuario	Positivo (%)	Neutral (%)	Negativo (%)
@themasterxtc	2,18 %	15,67 %	82,15 %
@silbersam	10,47 %	55,91 %	33,62 %
@Chagon16	3,52 %	18,43 %	78,05 %
@juanriv42678873	3,17 %	20,68 %	76,15 %
@Votamos__Todos	8,98 %	71,88 %	19,14 %
@AlejandraLisbo6	4,01 %	14,71 %	81,28 %

Tabla 8: 6 usuarios que más *tweets* publicaron sobre la convención constituyente entre enero y septiembre del 2022, categorizados por sentimiento.

Fuente: Elaboración propia.

Los resultados obtenidos en la Tabla 8 comprueban todas las hipótesis planteadas con anterioridad. Los 4 perfiles más cercanos a la derecha política promediaron un 79,4 % de *tweets* categorizados con una percepción negativa de la convención constituyente, notablemente más alto que los promedios de negatividad establecidos previamente (75,4 % en *hashtags*, 67,52 % en clasificación total y 62,64 % en negatividad promedio), mientras que los dos usuarios restantes se mostraron más neutrales frente al proceso aunque tampoco lo apoyaron, así como nadie lo hizo al menos en lo obtenido en dicha tabla. En síntesis, el porcentaje de *tweets* clasificados como positivos dentro del top 6 de usuarios que más *twittearon* sobre la convención fue de 5,39 %, los neutrales un 32,88 % y los negativos un 61,73 %, siguiendo la tendencia que hasta ahora ha parecido marcar todo el debate constitucional.

Si bien los resultados obtenidos en el análisis anterior son valiosos para la investigación, estudiar a aquellos usuarios que más fueron mencionados durante el debate resulta ser de mayor valor, debido a que a través de esto, es posible determinar un grado de influencia de éstos sobre las demás personas o del sentimiento que despertaron ellos sobre el resto. Nuevamente, se iteró sobre el *dataset* para encontrar dicha información y comenzar a estudiarla.

Usuario	Cantidad de menciones
@tere_marinovic	6.585
@gabrielboric	6.245
@biobio	4.520
@Jaime_Bassa	4.487
@latercera	3.746
@T13	3.501
@Mister_Wolf_0	3.055
@patriciapolitz	2.884
@mcubillosgall	2.830
@CNNChile	2.441

Tabla 9: 10 usuarios más mencionados en *tweets* sobre la convención constituyente entre enero y septiembre del 2022.

Fuente: Elaboración propia.

Como se puede ver en la Tabla 9, la mayoría de los perfiles hallados aquí corresponden a personalidades políticas importantes o a medios de comunicación masivos, dado que los usuarios más mencionados deben ser personas o instituciones que en la vida real tengan un gran alcance y sean conocidos por una parte importante de la población. En el listado se tiene al actual presidente de la república, Gabriel Boric (@gabrielboric), ex constituyentes que pertenecieron a la convención, como Teresa Marinovic (@tere_marinovic) y Jaime Bassa (@Jaime_Bassa), y medios de comunicación masivos como Radio Bío Bío (@biobio) y el Diario La Tercera (@latercera). De esta manera, fue posible hacer una subclasificación en los usuarios más mencionados entre:

- Constituyentes
- Personajes públicos
- Medios de comunicación

Sin embargo, dentro de los usuarios detallados en la Tabla 9, existe un *outlier*, @Mister_Wolf_0, cuya identidad no salta rápidamente a la vista. Es por esto que, al igual que como se hizo con los usuarios que más *twittearon* en el período constitucional, también se realizó un análisis de sus perfiles, pero esta vez tomando un *top 10* para cada subcategoría definida recientemente.

De esta manera, el primer grupo de usuarios a analizar fue nada menos que los protagonistas de este proceso, los constituyentes, cuyas visiones pueden haber influenciado a muchas personas durante su trayecto, como también sus propias acciones pueden haber sido blanco de críticas que se tradujeron en una mala percepción de la convención como organismo. En la Tabla 10, se listan los 10 constituyentes que más menciones obtuvieron a través del lapso estudiado, que luego fueron analizados individualmente sus perfiles, y finalmente se categorizaron por sentimiento, utilizando como base los *tweets* en los cuales fueron mencionados.

Usuario	Cantidad de menciones
@tere_marinovic	6.585
@Jaime_Bassa	4.487
@patriciapolitz	2.884
@mcubillosigall	2.830
@baradit	2.007
@fernando_atria	1.846
@RenatoGarinG	1.758
@IgnacioAchurra	1.449
@berfontaine	1.433
@arturozunigaj	1.276

Tabla 10: 10 constituyentes más mencionados en *tweets* sobre la convención constituyente entre enero y septiembre del 2022.

Fuente: Elaboración propia.

- Teresa Marinovic (@tere_marinovic): Licenciada en filosofía, fue electa como primera de su lista, Vamos Por Chile, en el distrito 10. Es cercana al partido Republicano, del ultraderechista José Antonio Kast y se ha declarado detractora de los gobiernos de Michelle Bachelet y el segundo mandato de Sebastián Piñera.
- Jaime Bassa (@Jaime_Bassa): Abogado constitucionalista de profesión, fue electo constituyente bajo la bandera de Apruebo Dignidad, coalición de izquierda, en el distrito 7 por la región de Valparaíso. En julio de 2021, fecha en que inició operaciones la Convención Constituyente, Bassa asumió el cargo de vicepresidente.
- Patricia Politzer (@patriciapolitz): Periodista, escritora y política, formó parte de la convención tras ser electa en el distrito 10, representando a la lista de los Independientes No Neutrales, personajes de la sociedad civil proclives al Apruebo. Fue presidenta del Consejo Nacional de Televisión durante los años 2001 y 2006, bajo el mandato de Ricardo Lagos y Michelle Bachelet.
- Marcela Cubillos (@mcubillosigall): Con una larga trayectoria política ligada a la UDI, partido de derecha, Cubillos fue diputada y luego ministra durante el segundo gobierno de Sebastián Piñera. Pese a declararse en contra del proceso constitucional, se postuló como constituyente por el distrito 11, finalmente obteniendo un escaño.

- Jorge Baradit (@baradit): Escritor y político, famoso por sus labores en divulgación histórica en medios de comunicación, así como también por su saga de libros titulados “Historia Secreta de Chile”. Ocupando un cupo de independiente por el Partido Socialista, Baradit fue escogido como constituyente por el distrito 10, luego formando parte del Colectivo Socialista, al cual finalmente renunció para unirse a las filas del Frente Amplio, coalición oficialista.
- Fernando Atria (@fernando_atria): Abogado, académico y político, militó durante 9 años en el Partido Socialista, postulándose como diputado y luego como presidente, sin tener éxito en ningún proceso. Tras desligarse del partido, fundó un nuevo movimiento llamado Fuerza Común, el cual no pudo constituirse como partido debido a la falta de afiliados. Sin embargo, tras recibir apoyo del pacto Apruebo Dignidad, Atria postuló como constituyente en el distrito 10, obteniendo un escaño con el 12,3% de los votos, primero de todo el distrito. Finalmente, en 2022, Fuerza Común se uniría al partido Convergencia Social, convirtiéndose Atria en militante de este último.
- Renato Garín (@RenatoGarinG): Ex militante de Revolución Democrática, partido perteneciente a la coalición oficialista, Frente Amplio, renunció a su puesto como diputado por el distrito 14 para lanzarse como candidato a la constituyente como independiente bajo apoyo del Partido Radical, obteniendo el tercer escaño por el mismo distrito.
- Ignacio Achurra (@IgnacioAchurra): Actor y dramaturgo, siguió los pasos políticos de su padre, Patricio Achurra, postulándose como constituyente por el distrito 14, bajo la militancia en su partido Convergencia Social, obteniendo el primer lugar en dicho distrito.
- Bernardo Fontaine (@berfontaine): Ingeniero comercial, que ganó notoriedad por ser un fuerte crítico de la reforma tributaria impulsada por la ex presidenta Michelle Bachelet en el 2014. En el 2021, se postularía como constituyente bajo la lista Vamos Por Chile, obteniendo el quinto escaño en el distrito 11. Durante el proceso constituyente, estuvo afiliado al movimiento Con Mi Plata No, iniciativa constitucional que pese a contar con la primera mayoría de apoyos, no fue considerada en el borrador por la convención, despertando varias críticas hacia dicho organismo.
- Arturo Zúñiga (@arturozunigaj): Ex subsecretario de redes asistenciales durante el segundo período de Sebastián Piñera, es un ingeniero comercial ligado a la UDI, que bajo dicha bandera postuló como constituyente por el distrito 9, obteniendo el quinto lugar. En la convención, fue sancionado por la Comisión de Ética tras dichos polémicos en contra de sus compañeros Marcos Barraza y Francisca Linconao.

El análisis de los perfiles de los 10 constituyentes recientemente presentados entrega una nueva forma de mirar el debate constitucional, dado que personajes ligados a la izquierda chilena, que en su mayoría se declararon a favor del Apruebo, por primera vez aparecieron en este estudio. Sin embargo, sus apariciones se ven opacadas por el poder mediático de

la constituyente de ultraderecha, Teresa Marinovic, que obtuvo la mayor cantidad de menciones, 6.585, 2.098 más que su seguidor más cercano, Jaime Bassa, quien fue el primer vicepresidente del organismo. En resumen, al categorizar por posición en cuanto al proceso, se puede decir que:

- Proclives al Apruebo (Jaime Bassa, Patricia Politzer, Jorge Baradit, Fernando Atria, Renato Garín, Ignacio Achurra): 14.431 menciones
- Proclives al Rechazo (Teresa Marinovic, Marcela Cubillos, Bernardo Fontaine, Arturo Zúñiga): 12.124 menciones

Teniendo por primera vez una mayoría que favorece, en primera instancia, al bando del Apruebo, que finalmente terminaría perdiendo, se puede decir que se presenta una nueva arista desde donde mirar la convención constituyente. Sin embargo, el simple hecho de ser mencionados sólo indica que son personajes de un gran poder mediático, pero no es posible determinar hacia qué lado fue su influencia, es decir, si sus acciones y dichos afectaron a la percepción del proceso de una manera positiva o negativa. Con el objetivo de determinar el grado de influencia de estos constituyentes, se procedió a analizar el sentimiento general contenido en cada uno de los tweets donde fueron mencionados.

Usuario	Positivo (%)	Neutral (%)	Negativo (%)
@tere_marinovic	7,79 %	14,03 %	78,18 %
@Jaime_Bassa	8,51 %	15,40 %	76,09 %
@patriciapolitz	4,40 %	13,25 %	82,35 %
@mcubillosgall	4,63 %	17,88 %	77,49 %
@baradit	9,92 %	17,64 %	72,45 %
@fernando_atria	3,85 %	18,96 %	77,19 %
@RenatoGarinG	6,20 %	22,58 %	71,22 %
@IgnacioAchurra	7,38 %	14,42 %	78,19 %
@berfontaine	6,21 %	15,49 %	78,30 %
@arturozunigaj	4,39 %	10,89 %	84,72 %

Tabla 11: 10 constituyentes más mencionados en tweets sobre la convención constituyente entre enero y septiembre del 2022, categorizados por sentimiento.

Fuente: Elaboración propia.

Siguiendo la tendencia que se fue marcando con cada vez más fuerza en cada análisis, las opiniones despertadas por estos constituyentes fueron mayoritariamente negativas, en promedio, mucho más altas que la negatividad detectada anteriormente, según lo mostrado en la Tabla 11, con un 77,6 %. También, en este análisis se encuentra el porcentaje de tweets clasificados como neutrales más bajo encontrado en este estudio, con un 16,1 %, mientras que la “positividad” se encuentra en un 6,3 %, cercana a los números previamente obtenidos.

Como era de esperar, el porcentaje de *tweets* negativos más altos encontrado correspondió a un constituyente de derecha, Arturo Zúñiga, con un 84,72 %, mientras que el nivel de negatividad más bajo lo obtuvo un constituyente de izquierda, Renato Garín, con 71,22 %. Sin embargo, un curioso hallazgo se encuentra al analizar los porcentajes de *tweets* positivos, el más alto corresponde a Jorge Baradit, con casi un 10 %, pero el más bajo corresponde también a un personaje proclive a la izquierda y al Apruebo, Fernando Atria, que alcanzó sólo un 3,85 % de positividad. La interpretación de dichas cifras puede deberse a múltiples aspectos, en el caso de los constituyentes apoyados por partidos de derecha, su labor fue ampliamente criticada tras ser acusados de intentar “entorpecer” el desarrollo de la convención constituyente, teniendo en cuenta también que muchos de estos personajes se manifestaron en contra de la mera existencia del proceso. Para los constituyentes proclives a la izquierda chilena, existen puntos de inflexión que fueron estudiados más adelante y que logran explicar el por qué fallaron en influenciar de manera positiva a la percepción del proceso constituyente; sin embargo, por ahora es correcto decir que influenciaron más negativa que positivamente.

Con esto, se procedió a investigar acerca de aquellos personajes con alto poder mediático que no fueron constituyentes pero que sí tuvieron un rol importante en el debate constituyente, ya sea porque no se postularon o porque no lograron ser parte, que se pueden ver en la Tabla 12.

Usuario	Cantidad de menciones
@gabrielboric	6.245
@Mister_Wolf_0	3.055
@mmlagoscc	2.096
@ximerincon	1.678
@joseantoniokast	1.432
@melnicksergio	1.376
@jschaulsohn	1.368
@jgalemparte	1.289
@felipekast	1.137
@MauricioMMQ75	907

Tabla 12: 10 no constituyentes más mencionados en *tweets* sobre la convención constituyente entre enero y septiembre del 2022.

Fuente: Elaboración propia.

Algunos de estos nombres saltan inmediatamente a la vista, como el presidente Gabriel Boric (@gabrielboric) y su contendiente en las elecciones presidenciales del 2021, José Antonio Kast (@joseantoniokast), sin embargo, existen otros personajes que valieron la pena estudiar y analizar desde el punto de vista de sus perfiles políticos, para tener una primera idea de su posición respecto al proceso constitucional.

- Gabriel Boric (@gabrielboric): Hijo de migrantes croatas, es el actual presidente de Chile tras vencer a José Antonio Kast en la segunda vuelta de las elecciones presidenciales

del 2021. En su trayectoria política ha sido dirigente estudiantil y diputado, siempre ligado a la izquierda, participando de la creación de la coalición que hoy lo apoya, el Frente Amplio. Antes de postularse a la presidencia, Boric fue parte del acuerdo que originó el inicio del proceso constitucional en Chile, declarándose a favor de este. Durante su mandato, muchas veces fue acusado por la oposición de intentar influir en el proceso, sosteniendo que su programa político se vería condicionado si es que el Rechazo ganaba.

- Mr. Wolf (@Mister_Wolf_0): Es el pseudónimo utilizado por una persona anónima que creó esta cuenta en Twitter luego del estallido social del 2019, para *twittear* acerca de las marchas y los sucesos relacionados, mostrándose a favor de éstas, convirtiéndose un tiempo después en columnista para el medio independiente Interferencia. Para la campaña presidencial del 2021, Mr. Wolf le dio su apoyo al actual presidente Gabriel Boric y se vio envuelto en ásperos debates con los simpatizantes de José Antonio Kast, quienes llegaron a amenazarlo de revelar su identidad real.
- Marta Lagos (@mmlagoscc): Encuestadora y fundadora de MORI Chile (*Market Opinion Research International*) y Latinobarómetro, es una *twittera* muy activa que suele comentar sobre política chilena. Se manifiesta contraria al golpe militar de 1973 y al actual proceso constituyente de 2023, cuya mayoría de integrantes pertenecen a la derecha.
- Ximena Rincón (@ximerincon): Abogada y política con una amplia trayectoria ligada a la centroizquierda, de la mano de su ex partido, la Democracia Cristiana, Rincón fue intendenta, senadora, ministra y precandidata presidencial. Ella, junto con su partido, se manifestó a favor del proceso constituyente y del Apruebo, sin embargo, cerca de la fecha del plebiscito, Rincón, junto con otros compañeros de bancada, restarían el apoyo al Apruebo, sumándose al Rechazo y formando parte de campañas por esta opción, como “Una que nos una” y “Centroizquierda x el Rechazo”. Finalmente, formaría un nuevo partido político llamado Demócratas.
- José Antonio Kast (@joseantoniokast): Abogado y político de ascendencia alemana, cuenta con una gran trayectoria política ligada a la derecha en su paso por la UDI (Unión Demócrata Independiente) y luego en la ultraderecha con su partido Republicano. Luego de ser concejal por la comuna de Buin y después diputado por el distrito 30 y 24, Kast impulsó su carrera al postularse a la carrera presidencial del 2017 como independiente, quedando cuarto en primera vuelta con un 8 % de los votos. Cuatro años más tarde, en 2021, volvería a postularse para presidente de la república, esta vez bajo la bandera del Partido Republicano, ganando la primera vuelta con un 27,91 % de los votos, pero perdiendo en segunda vuelta contra el actual presidente Gabriel Boric. Fue un acérrimo detractor del proceso constitucional y se ha declarado a favor del golpe militar de 1973, siendo seguidor del dictador Augusto Pinochet, a quien apoyó en la franja del Sí para el plebiscito de 1988.
- Sergio Melnick (@melnicksergio): Economista, escritor y político, formó parte de la dictadura militar de Augusto Pinochet como ministro de Estado. También ha sido gerente

general de varias empresas y director ejecutivo de dos canales de televisión, actualmente, se desempeña como concejal por la comuna de Las Condes. Declarado a sí mismo como pinochetista y cercano al partido de José Antonio Kast, también fue un personaje que se mostró contrario al proceso constitucional desde su inicio, defendiendo la Constitución de 1980, escrita durante la dictadura.

- Jorge Schaulsohn (@jschaulsohn): Abogado y político, fue militante del Partido Radical para luego participar en la creación del Partido Por la Democracia (PPD). Fue diputado por el distrito 22 y presidió la cámara durante el año 1994. Schaulsohn decidió postularse como independiente a la convención constituyente, declarando no sentirse representado por ningún partido político, especialmente por la izquierda. Sin embargo, no logró reunir los patrocinios necesarios para lanzar su candidatura y quedó fuera del proceso.
- Gabriel Alemparte (@jgalemparte): Abogado y cientista político, fue militante por casi 20 años de la Democracia Cristiana, al cual renunció en 2020. Abogó por la existencia de un proceso constitucional, al cual apoyó hasta el lanzamiento del borrador constitucional, momento en el cual su visión cambió, llegando a formar parte del movimiento “Centroizquierda x el Rechazo”. Actualmente se desempeña como el vicepresidente del partido de Ximena Rincón, Demócratas.
- Felipe Kast (@felipekast): Fundador del partido de derecha, Evópoli, es sobrino del ultraderechista José Antonio Kast, y a lo largo de su carrera política, se ha desempeñado como diputado, senador y candidato presidencial. Se mostró contrario al proceso constitucional, difundiendo múltiples *fake news* para desprestigiar a la convención y su trabajo.
- Mauricio Morales (@MauricioMMQ75): Analista político y académico cercano a la Democracia Cristiana, *twitea* activamente sobre política chilena desde un tono moderado, criticando a la izquierda y a la derecha chilena. También criticó el borrador constitucional, asegurando que estaba redactado en clave de los 90 y no de los 2020.

Al igual que el grupo de los ex constituyentes, este conjunto de personas es bastante heterogéneo en cuanto a posiciones políticas, tanto de izquierda, centro y derecha. Llama particularmente la atención la presencia de personajes que cambiaron su opinión durante el proceso, pasando de apoyar el Apruebo a llamando a votar por el Rechazo, incluso formando parte de campañas a favor de esta última opción, como Ximena Rincón y Gabriel Alemparte, que formaron parte del movimiento “Centroizquierda x el Rechazo” y que ahora forman parte del nuevo partido político Demócratas. También es importante destacar la presencia de personajes declarados “pinochetistas”, que se opusieron rotundamente a la existencia de un proceso constitucional, dado que reemplazaría la Constitución de 1980, escrita durante la dictadura militar.

Dentro de este grupo, el presidente Gabriel Boric es el personaje con más menciones, lo cual se entiende en base a que antes de convertirse en mandatario fue uno de los impulsores

principales de este proceso, el cual siguió atentamente desde La Moneda, hecho que le valió varias críticas desde la derecha en todo su espectro, ya que lo acusaron de “favorecer” al Apruebo. Boric, junto con Mr. Wolf, el segundo más mencionado, son las únicas personalidades que se declaran abiertamente de izquierda dentro de este top 10. Así, fue posible hacer la siguiente separación:

- Proclives al Apruebo (Gabriel Boric, Mr.Wolf): 9.300 menciones
- Neutrales/No se sabe (Marta Lagos, Jorge Schaulsohn, Mauricio Morales): 4.371 menciones
- Proclives al Rechazo (Ximena Rincón, José Antonio Kast, Sergio Melnick, Gabriel Alemarte, Felipe Kast): 6.912 menciones

Sin embargo, y al igual que con el ejercicio pasado, se analizó el sentimiento general de los *tweets* desde los cuales fueron mencionados estos personajes, con el fin de determinar si su presencia afectó positiva o negativamente a la percepción de los usuarios de Twitter respecto de la convención constituyente, lo que se resume en la Tabla 13.

Usuario	Positivo (%)	Neutral(%)	Negativo (%)
@gabrielboric	6,08 %	20,22 %	73,69 %
@Mister_Wolf_0	9,39 %	17,91 %	72,70 %
@mmlagoscc	3,77 %	21,28 %	74,95 %
@ximerincon	7,63 %	20,44 %	71,93 %
@joseantoniokast	3,42 %	15,64 %	80,94 %
@melnicksergio	2,54 %	15,70 %	81,76 %
@jschaulsohn	3,73 %	24,56 %	71,71 %
@jgalemparte	4,73 %	24,44 %	70,83 %
@felipekast	3,25 %	17,33 %	79,42 %
@MauricioMMQ75	5,62 %	28,11 %	66,26 %

Tabla 13: 10 no constituyentes más mencionados en *tweets* sobre la convención constituyente entre enero y septiembre del 2022, categorizados por sentimiento.

Fuente: Elaboración propia.

Nuevamente, las percepciones en general son bastante negativas, confirmando la tendencia de todos los análisis realizados hasta este punto. Sin embargo, es importante destacar que los tres personajes que más se opusieron al proceso, José Antonio Kast, Sergio Melnick y Felipe Kast, efectivamente lograron influenciar de manera negativa el parecer de los usuarios de Twitter respecto del proceso, dado que presentaron los niveles más altos de negatividad en este apartado (80,94 %, 81,76 % y 79,42 %, respectivamente), así como también lograron los niveles de positividad más bajos: 3,42 %, 2,54 % y 3,25 %. Por su parte, el *twittero* anónimo Mr. Wolf hizo una mejor labor en intentar sumar apoyos por el Apruebo, que de todas

maneras quedó bastante corto, alcanzando solamente un 9,39 %, el más alto del grupo. Otro aspecto importante a mencionar es el hecho de que los personajes que fueron catalogados como “neutrales” en un inicio mantuvieron su “neutralidad”, debido a que presentaron, en promedio, la mayor cantidad de tweets catalogados como neutrales, con un 24,65 % combinado entre Lagos, Schaulsohn y Morales. Por último, el desempeño de los “conversos”, Rincón y Alemparte, estuvo en los niveles esperados, concentraron una gran cantidad de negatividad, pero también una cantidad considerable de neutralidad, teniendo que Alemparte, con un 24,44 %, fue el tercero más neutral de este grupo, lo cual encuentra su explicación en el hecho de que ambos personajes estuvieron a favor en un inicio, generando mayor cantidad de reacciones positivas y neutrales en un comienzo, que luego se fueron transformando en negativas conforme a su postura cambiaba en 180 grados.

Finalmente, el último análisis de esta sección se enfocó en el rol de los medios de comunicación en el debate constitucional, el cual es importante observar y analizar debido al gran poder mediático que éstos poseen y que logran influir de gran manera en la opinión de las personas. En la Tabla 14, se presentan los 10 medios de comunicación más mencionados en Twitter durante enero y septiembre de 2022.

Usuario	Cantidad de menciones
Radio Bío Bío (@biobio)	4.520
Diario La Tercera (@latercera)	3.746
Teletrece (@T13)	3.501
CNN Chile (@CNNChile)	2.441
ADN Radio (@adnradiochile)	2.373
Radio Cooperativa (@Cooperativa)	2.084
Mega Noticias (@meganoticiascl)	1.293
Chilevisión Noticias (@CHVNoticias)	1.280
The Clinic (@thecliniccl)	1.131
24 Horas (@24HorasTVN)	1.020

Tabla 14: 10 medios de comunicación más mencionados en tweets sobre la convención constituyente entre enero y septiembre del 2022.

Fuente: Elaboración propia.

A diferencia de los dos grupos estudiados anteriormente, no se realizó un análisis de perfiles para cada medio identificado, dado que el análisis de su línea editorial puede ser confusa y principalmente tienen roles de difusión en este tipo de procesos. Analizar sesgo dentro de las noticias emitidas por estos medios es una tarea más compleja que requeriría de un nuevo estudio completo que se encargue de eso. En su lugar, se categorizaron por el tipo de prensa que representan:

- Prensa escrita (La Tercera, The Clinic): 4.877 menciones
- Prensa radial (Bío Bío, ADN, Cooperativa): 8.977 menciones

- Prensa audiovisual/digital (Teletrece, CNN Chile, Mega Noticias, Chilevisión Noticias, 24 Horas): 9.535 menciones

Dado que el origen de los datos estudiados es Twitter, resulta natural que los medios digitales concentren la mayor cantidad de menciones, aunque sorprendentemente los medios radiales les siguieron muy de cerca con una diferencia de apenas 500 menciones. En contraparte, la prensa escrita se queda muy por detrás con apenas 4.877 menciones y sólo dos representantes, de los cuales sólo uno es un medio escrito tradicional, La Tercera, dado que The Clinic es un medio con un tono más satírico que se publica semanalmente, aunque tiene una gran presencia en redes sociales, como se ve demostrado aquí. Ahora bien, el foco del análisis tiene que ver con el sentimiento contenido en los *tweets* en donde se vieron mencionados estos medios, los cuales se presentan en la Tabla 15. Los números presentados muestran los niveles de negatividad más altos encontrados en todo el estudio hasta ahora, con un 77,89 % de *tweets* detectados con una percepción negativa, sin embargo, es importante señalar que esta cifra no significa que los medios hayan hecho una campaña en contra del Apruebo y el proceso constitucional, sino que las personas que leyeron las publicaciones de estos medios, habitualmente noticias, reaccionaron en contra de los sucesos ocurridos, los cuales fueron analizados más adelante. En efecto, se comprueba que la percepción de los usuarios de Twitter sobre la convención constituyente en general fue muy mala, y esta negatividad se acrecentó al consumir noticias sobre el quehacer de dicho organismo.

Usuario	Positivo (%)	Neutral (%)	Negativo (%)
@biobio	3,19 %	17,61 %	79,20 %
@latercera	3,55 %	19,11 %	77,34 %
@T13	3,40 %	15,85 %	80,75 %
@CNNChile	5,16 %	19,75 %	75,09 %
@adnradiochile	3,75 %	18,08 %	78,17 %
@Cooperativa	5,47 %	20,73 %	73,80 %
@meganoticiascl	3,02 %	17,17 %	79,81 %
@CHVNoticias	3,05 %	17,73 %	79,22 %
@thecliniccl	4,33 %	17,60 %	78,07 %
@24HorasTVN	6,96 %	15,59 %	77,45 %

Tabla 15: 10 medios más mencionados en *tweets* sobre la convención constituyente entre enero y septiembre del 2022, categorizados por sentimiento.

Fuente: Elaboración propia.

Con esto, se dio por finalizada la primera etapa y más extensa del análisis basado en Py-Sentimiento, el análisis de sentimiento, el cual concluyó que la percepción de la gente en Twitter fue bastante negativa respecto de la convención constituyente, lo cual se comprobó mediante análisis generales que después se fueron especificando, pasando por los *hashtags* más recurrentes y, finalmente, por los usuarios y sus características. En el apartado siguiente, una nueva dimensión se añade al estudio de los hechos y tendencias que determinaron

el resultado del plebiscito constitucional de 2022, mediante el estudio de las emociones detectadas en los *tweets* extraídos.

3.5.2. Análisis de emociones

Estudiar los sentimientos contenidos en los *tweets* publicados por usuarios que escribieron sobre la convención constituyente abrió una nueva ventana en la forma de entender el proceso constituyente, entendiendo la percepción que éstos tenían de un organismo nuevo cuya misión era redactar una nueva carta magna, tarea complicada entendiendo el panorama político en Chile durante los últimos 50 años. No obstante, profundizar el estudio del lenguaje natural empleado en cada *tweet* siguió ampliando el conocimiento acerca del pensamiento de las personas que finalmente decidieron rechazar la propuesta constitucional redactada por dicho organismo. Esta sección del análisis mediante PySentimiento se enfocó en las emociones expresadas por los *twitteros* al momento de hablar sobre la convención constituyente, con el fin de ahondar y dar nuevas perspectivas sobre las tendencias ya identificadas a lo largo de este documento.

Antes de continuar, es importante señalar que la estrategia de “lo general a lo particular” se siguió adoptando, sólo que en un menor grado de profundidad para que no se alargara innecesariamente este trabajo. Dicho esto, se presentan los tiempos de procesamiento empleados para cada *subdataset* y la cantidad de *tweets* analizados en la Tabla 16.

Previo al análisis en cuanto al tiempo y la cantidad de *tweets* procesados, es importante acotar que durante el desarrollo de este estudio, fue posible acceder a un nuevo equipo con mejores especificaciones, el cual mejoró drásticamente el rendimiento obtenido con las especificaciones anteriores. A partir de este momento, los componentes con los cuales se trabajó fueron los siguientes:

- CPU: AMD Ryzen 7 5800X, 8 núcleos y 16 hilos, frecuencia base de 3,8 GHz, alcanzando los 4,7 GHz en máxima carga
- Memoria RAM: 16 GB DDR4, dual-channel, 3600 MHz
- GPU: Nvidia Geforce RTX 3070 Ti, 8 GB GDDR6X 256-bit
- Almacenamiento: 1 TB M.2 NVMe PCIe 4.0, 5000 MB/s en lectura, 4200 MB/s en escritura

Mes	Cantidad de tweets	Tiempo de procesamiento
Enero	35.346	28:21
Febrero	44.879	36:38
Marzo	39.902	34:54
Abril	32.160	26:59
Mayo	43.496	38:56
Junio	39.070	36:11
Julio	43.412	07:42
Agosto	42.623	08:06

Tabla 16: Volumen del *dataset* con respuestas y su tiempo de procesamiento para análisis de emociones.

Fuente: Elaboración Propia.

Con un equipo de mejores especificaciones y con acceso a las bibliotecas y *drivers* para hacer uso de los núcleos CUDA presentes en la tarjeta gráfica, se aumentó considerablemente la velocidad de procesamiento de *tweets*, pasando de una velocidad promedio de 19,4 *tweets* por segundo a una de 90,8 *tweets* por segundo, es decir, un rendimiento aproximadamente 4,5 veces mejor. Es importante recalcar que esta mejora en rendimiento se debe gracias al acceso a una mejor tarjeta gráfica, ya que si se comparan los procesadores de ambos equipos no existen diferencias significativas, es más, al realizarse pruebas en base al procesamiento por CPU, los tiempos de ejecución fueron similares.

Una vez terminado el estudio, se obtuvieron las emociones predominantes por cada *tweet*, las cuales se presentan en la Tabla 17. Previo a la interpretación de los resultados, es importante señalar que para este análisis, muchas veces PySentimiento, con la ayuda de EmoEval, no fue capaz de identificar una emoción predominante en un *tweet*, pasando a clasificarse como “neutral” en emoción, lo cual, para propósitos de este estudio, no aporta mayor información, por lo que no fue considerado. A su vez, las emociones identificables mediante esta metodología son las siguientes: tristeza, enojo, disgusto, miedo, sorpresa y alegría, de los cuales sólo tres aparecen en la Tabla 17. Esto se debe a que el disgusto, el miedo y la sorpresa no alcanzaron un porcentaje mayor al 1% en ningún mes contenido en el *dataset*, por lo que fueron considerados irrelevantes.

Mes	Tristeza (%)	Enojo (%)	Alegría (%)
Enero	2,37 %	31,69 %	4,30 %
Febrero	2,02 %	43,03 %	2,14 %
Marzo	2,28 %	39,12 %	2,35 %
Abril	2,27 %	42,01 %	1,83 %
Mayo	2,12 %	41,08 %	2,31 %
Junio	2,23 %	38,02 %	2,25 %
Julio	1,78 %	36,68 %	2,19 %
Agosto	2,51 %	38,29 %	2,28 %

Tabla 17: Análisis de emociones aplicado sobre el dataset con respuestas
Fuente: Elaboración propia.

Ya hechas las acotaciones, el análisis de las emociones encontradas en el *dataset* con respuestas se condicen en gran parte con las tendencias establecidas en el análisis de sentimiento, debido a que tanto el enojo como la tristeza, consideradas emociones negativas, son las que predominan en los resultados obtenidos, dejando a la alegría como la única emoción positiva identificada, con un porcentaje mucho menor respecto a estas dos, al igual que la positividad en los análisis previamente desarrollados. El enojo fue la emoción más detectada, promediando un 38,74 % de tweets que expresaron dicha emoción al hablar de la convención constituyente; le siguen la alegría con un 2,46 % y la tristeza con un 2,2 %. En cuanto a distribución, es importante señalar una tendencia que se fue marcando silenciosamente durante este trabajo y que no ha sido analizada a fondo; si bien la tristeza presenta una distribución bastante uniforme, al igual que la alegría a excepción del mes de enero, puede notarse que al mes siguiente, febrero, el enojo sufre una drástica subida, de 31,69 % a 43,03 %, es decir, casi 12 puntos porcentuales más, alza que se mantuvo entre los meses de febrero y mayo, hasta finalmente comenzar a disminuir en junio del 2022. Los factores y sucesos que expliquen dicho movimiento serán analizados en la sección posterior, titulada “Detección de discurso de odio”.

Con un análisis general ya realizado, descubriendo que el enojo fue la emoción predominante durante el debate constitucional y que tuvo una inusitada, por ahora, alza entre los meses de febrero y mayo, se atacaron aspectos más particulares con los cuales fue posible obtener más información acerca del objeto de estudio de este trabajo, enfocándose nuevamente en los *hashtags*, con el objetivo de identificar y entender las emociones encontradas en los tweets que los utilizaron.

Hashtag	Emoción	Apariciones
#Rechazo	Enojo	956
#RechazoDeSalida2022	Enojo	586
#RechazoTransversal	Enojo	586
#ConvencionCulia	Enojo	508
#RechazoElMamarrachoComunista	Enojo	472
#RechazoDeSalida	Enojo	455
#RechazoElMamarracho	Enojo	429
#ConvencionConstitucional	Enojo	427
#Apruebo	Alegría	368
#RechazoLaDestruccionDeChile	Enojo	246

Tabla 18: 10 *Hashtags* más recurrentes encontrados en la clasificación por emoción.

Fuente: Elaboración propia.

Al igual que en el análisis general, el enojo nuevamente es la emoción predominante en el análisis, como se puede ver en la Tabla 18, con 9 de los 10 *hashtags* más utilizados proviniendo desde *tweets* en los cuales predominaba el enojo como emoción principal. El *hashtag* restante, ubicado en el noveno lugar, proviene de la segunda emoción más predominante, la alegría, a través del *hashtag* #Apruebo, que en análisis anteriores ya había sido calificado como un *hashtag* de connotación positiva hacia la convención constituyente, mientras que el resto de los *hashtags* que están relacionados al enojo fueron clasificados antes con un sentimiento negativo, lo cual permite establecer una relación causal entre el enojo y el sentimiento negativo, y la alegría con el sentimiento positivo, condiciéndose con la realidad de que aquellas personas que apoyaban el trabajo de la convención constituyente, estaban de acuerdo con votar por el Apruebo.

De esta manera, concluye el análisis de emociones aplicado a los *tweets* que hablan sobre la convención constituyente en el lapso comprendido entre enero y septiembre del 2022, realizando importantes hallazgos acerca del parecer de la gente respecto del proceso constituyente, identificando en su mayoría una sensación de enojo al momento de hablar del tema, lo cual se tradujo en un sentimiento negativo mediante el análisis de los *hashtags*, que establecieron una relación directa entre el enojo y el Rechazo, y la alegría con el Apruebo. Finalmente, quedó propuesto para la siguiente sección del análisis el identificar la creciente ola de rechazo y enojo que se dio hacia el trabajo de la convención constituyente entre los meses de febrero y mayo del 2022.

3.5.3. Detección de discurso de odio

A lo largo de este estudio, se aplicaron diversas técnicas y utilizado múltiples modelos para visualizar el panorama constitucional desde un punto de vista más humano, mediante sus emociones y sentimientos. Los resultados hasta ahora indican que la negatividad respecto del

trabajo de la convención constituyente predomina por sobre la neutralidad, y por sobre todo, la positividad, dando a entender que los usuarios de Twitter jamás estuvieron en su mayoría de acuerdo con el trabajo y lo propuesto por este organismo. Apoyando a la tendencia ya identificada, el análisis de emociones apunta a que el enojo fue la emoción más común entre los *twitteros*, lo cual se traduce en un rechazo inmediato hacia el trabajo constitucional que se realizó en Chile durante en el 2022, confirmándose a través de la relación causal que se identificó entre el sentimiento negativo y el enojo, mediante el análisis específico aplicado a los *hashtags* más populares durante el período.

Todos estos antecedentes fueron puestos a prueba una vez más a través de la detección del discurso de odio, gracias a PySentimiento y al modelo especializado SemEval, que permitió añadir las últimas aristas que componen el panorama que se ha construido con los análisis realizados anteriormente. Empleando la misma metodología que se ha utilizado hasta ahora, se aplicó detección de discurso de odio a todo el *dataset* presente, haciendo la distinción por meses, con el fin de obtener la mirada más general posible. Los resultados de dicho análisis se presentan en la tabla 19.

Mes	Discurso de odio (%)
Enero	13,08 %
Febrero	20,52 %
Marzo	16,94 %
Abril	16,97 %
Mayo	16,70 %
Junio	15,61 %
Julio	13,12 %
Agosto	13,14 %

Tabla 19: Detección de discurso de odio aplicada sobre el dataset con respuestas
Fuente: Elaboración propia.

Antes del análisis de los resultados obtenidos, es importante precisar que el modelo aplicado entrega más valores que sólo el porcentaje de odio obtenido en un *tweet*, sino que también puede detectar el nivel de agresividad presente en uno, como también el nivel de “direccionamiento” hacia algo o alguna persona. Sin embargo, en la Tabla 19 sólo se presenta el porcentaje de odio identificado, y esto es porque en ningún mes ninguna de las otras métricas superaron el 1 %, por lo cual fueron descartadas del estudio.

Entrando al análisis, se puede apreciar de que los niveles de discurso de odio se ven relativamente “bajos”, si se comparan con los niveles de negatividad y de enojo presentados anteriormente; aunque hace sentido, dado que un *tweet* puede tener una percepción negativa sobre la convención constituyente y además expresar enojo respecto al tema, pero no necesariamente expresar odio hacia el trabajo de dicho organismo. Por ende, se puede afirmar que los niveles de discurso de odio presentes fueron consistentes con los análisis realizados previos a este.

Más que los niveles de odio detectados y su volumen en cada mes, es llamativa la distribución que ésta presenta mes a mes, dado que se puede apreciar la misma curva creciente entre los meses de febrero y mayo que ya había sido detectada en la sección del análisis de emoción. Dado que esta tendencia volvió a aparecer, sugiere algún tipo de correlación entre los niveles de odio/enojo encontrados y algún suceso en particular ocurrido durante ese período, específicamente en el mes de febrero, lo cual fue estudiado con el fin de obtener más conclusiones acerca de los hechos y tendencias que finalmente marcaron el resultado del plebiscito constitucional del 2022.

Analizando los sucesos relacionados al proceso constitucional durante el mes de febrero destaca uno en particular ocurrido el 18 de dicho mes, la aprobación de la norma en la convención constitucional que estipulaba que Chile sería un “estado regional, plurinacional e intercultural”, alcanzando los 2/3 necesarios para ser aprobados pese a los intentos de los convencionales de derecha por rechazar dicha normativa y que no fuese incorporada en el borrador provisional de la presunta nueva Constitución de Chile. Dicho evento fue bastante bullado en redes sociales y en el debate constitucional en general, criticando principalmente la idea de considerar al pueblo mapuche dentro de la plurinacionalidad de Chile en base al conflicto mapuche que se ha mantenido vigente durante los últimos años en la región de la Araucanía, que en votaciones recientes se ha alineado con los partidos de derecha, quienes han tildado a los mapuches como “terroristas”. Teniendo en cuenta el potente trasfondo de este suceso, se planteó la hipótesis de que la aprobación de dicha norma provocó, en gran parte, el aumento en el odio y el enojo hacia la Convención Constitucional.

La metodología para comprobar o refutar dicha hipótesis fue simple y consistió de los siguientes pasos:

1. Elaborar un algoritmo que permita identificar los *tweets* que hablen acerca de la norma en cuestión, principalmente buscando por palabras claves como plurinacional e intercultural.
2. Extraer los *tweets* identificados por el algoritmo para su análisis.
3. Analizar la clasificación de odio obtenida para cada *tweet* identificado.
4. Tabular la presencia de odio por cada mes y verificar si existe alguna tendencia consistente con la inicial.

Siguiendo estos pasos, se elaboró un algoritmo el cual permitió extraer, por cada mes, la mayoría de los *tweets* relacionados con la norma referente a la plurinacionalidad e interculturalidad de Chile, tras lo cual fueron analizadas sus etiquetas, con el fin de comprobar si es que dichos *tweets* tenían presencia de discurso de odio en su interior.

Mes	Discurso de odio (%)
Enero	2,99 %
Febrero	23,70 %
Marzo	18,39 %
Abril	18,81 %
Mayo	15,78 %
Junio	9,59 %
Julio	4,65 %
Agosto	4,13 %

Tabla 20: Detección de discurso de odio en *tweets* referentes a Chile plurinacional e intercultural

Fuente: Elaboración propia.

Los porcentajes mostrados en la Tabla 20 corresponden al porcentaje total de *tweets* referentes a la norma constitucional que habla de Chile plurinacional e intercultural, en los cuales fue detectado discurso de odio respecto del total de *tweets* clasificados como “odiosos”. Para aclarar, se puede tomar el ejemplo del mes de febrero, en donde un 23,7 % de los *tweets* que contenían discurso de odio hablaban acerca de dicha norma. Así, podemos visualizar la misma tendencia que en análisis anteriores, ya que desde enero a febrero se nota un crecimiento de un 20,7 %, que a partir de marzo comienza lentamente a disminuir hasta normalizarse en el mes de junio. Lo que restaba por analizar, era el peso que realmente tuvieron estos porcentajes por sobre los obtenidos en la Tabla 19.

Mes	Odio sin plurinacionalidad (%)	Odio con plurinacionalidad (%)	Aumento (%)
Enero	12,69 %	13,03 %	0,39 %
Febrero	15,66 %	20,52 %	4,86 %
Marzo	13,82 %	16,94 %	3,12 %
Abril	13,78 %	16,97 %	3,19 %
Mayo	14,06 %	16,70 %	2,64 %
Junio	14,11 %	15,61 %	1,50 %
Julio	12,51 %	13,12 %	0,61 %
Agosto	12,60 %	13,14 %	0,54 %

Tabla 21: Detección del aumento en discurso de odio provocado por *tweets* referentes a Chile plurinacional e intercultural

Fuente: Elaboración propia.

Como se puede apreciar en la Tabla 21, queda de manifiesto que en febrero, el mes en que se aprobó la norma que declaraba que Chile sería una nación plurinacional e intercultural, la presencia de discurso de odio se elevó considerablemente, pasando de un 0,39 % en enero, mes en que se comenzó a discutir la norma, a un 4,86 % de aumento en febrero, mes en que

fue fuertemente discutida y finalmente aprobada. Luego, se puede ver cómo el incremento comienza a disminuir a partir de marzo, pero se sostiene durante abril y mayo, cayendo en junio para finalmente prácticamente desaparecer a partir de julio, meses en donde la presencia de estos *tweets* no supera el 1 %.

De esta manera, se comprobó la hipótesis de que la aprobación de la norma que declaraba a Chile como un país plurinacional e intercultural fue un evento que elevó considerablemente el odio y el enojo hacia la convención constituyente, aumentando en casi cinco puntos porcentuales la cantidad de *tweets* con presencia de discurso de odio hacia dicho organismo en tan sólo un mes, alza que si bien disminuyó paulatinamente, se mantuvo por aproximadamente cuatro meses y sentó una de las bases para el fracaso del borrador constitucional, que finalmente fue rechazado. Sin embargo, no es posible afirmar que la aprobación de dicha norma fue el único factor que causó el triunfo del Rechazo, dado que, si se aísla la presencia de este evento de los datos estudiados, aún así existe un alza sostenida que se mantuvo durante un lapso más amplio de tiempo y que se condice tanto con el análisis de emociones y de sentimiento, así como también con lo mostrado en la Definición del Problema y la tendencia que mostraban las encuestas, lo cual puede haberse visto asociada a otros sucesos o a un arrastre de hechos que ocurrieron cronológicamente antes que los sucesos estudiados. Como puede haber sido el caso del ex convencional Rodrigo Rojas Vade, quien admitió en una entrevista haber falsificado su diagnóstico de cáncer y el personaje que lo llevó a ser uno de los rostros del estallido social y, posteriormente, de la convención constitucional, causándole un daño irreparable a la imagen pública del proceso y sus participantes.

Ya identificadas las razones detrás de la tendencia más grande e importante encontrada a lo largo de esta investigación, se dio una vuelta más sobre los *hashtags* que se hicieron presentes durante el debate, esta vez desde el punto de la presencia del discurso de odio, con el objetivo de saber si era posible establecer una relación de causalidad entre los *hashtags* relativos al Rechazo y el odio hacia la convención constitucional, así como ya lo fueron en los análisis de sentimiento y emociones, en donde los *hashtags* pertenecientes al Rechazo se vieron directamente asociados con una sensación de negatividad y enojo por parte de los usuarios de Twitter que comentaron el proceso.

Hashtag	¿Discurso de odio?	Apariciones
#ConvencionCulia	Sí	286
#Rechazo	Sí	276
#RechazoDeSalida2022	Sí	251
#RechazoDeSalida	Sí	185
#RechazoTransversal	Sí	183
#ConvencionConstitucional	Sí	136
#RechazoElMamarrachoComunista	Sí	94
#ConvencionalesCSM	Sí	93
#Chile	Sí	75
#IzquierdaMiserable	Sí	61

Tabla 22: 10 *Hashtags* más recurrentes encontrados en la detección de discurso de odio.
Fuente: Elaboración propia.

Todos los *hashtags* presentados en la Table 22 corresponden a *tweets* que contienen una cantidad considerable de discurso de odio en su interior, de los cuales se pueden reconocer inmediatamente varios *hashtags* que ya estaban relacionados a un sentimiento negativo y/o al enojo en análisis anteriores, como lo son #ConvencionCulia, #Rechazo, y todos los *hashtags* que contienen en su interior la palabra “Rechazo”, avalando una relación causal más entre *hashtags* de connotación negativa y el Rechazo. Además de esto, es importante notar que en este análisis han aparecido *hashtags* que hasta ahora no se habían presentado en este estudio, como por ejemplo #ConvencionalesCSM, #Chile e #IzquierdaMiserable. El primer *hashtag* es un derivado del primer *hashtag* encontrado en esta lista, #ConvencionCulia, pero interpelando de manera odiosa y negativa directamente a los convencionales como individuos más que al organismo. Por su parte, #Chile hace referencia al nombre de nuestro país, por lo cual no fue posible obtener mayor información mediante su análisis, sin embargo, es posible que dicho *hashtag* haya estado relacionado directamente con la tendencia de “defender Chile” que adoptaron varios personajes de derecha y votantes del Rechazo con el fin de evitar el cambio de la actual constitución. Finalmente, #IzquierdaMiserable es un ataque directo a dicho sector político, el cual, como ya se ha hablado en el transcurso de este documento, fue en su mayoría asiduo al Apruebo y además fue la fuerza oficialista por la mayor parte del tiempo analizado en este *dataset*. De esta forma, fue posible establecer una correlación entre los *hashtags* provenientes de *tweets* con contenido de discurso de odio y el Rechazo.

Así, se dio por terminada toda la etapa referente al análisis de los *tweets* sobre la convención constitucional entre los meses de enero y septiembre de 2022, pasando por el análisis de sentimiento, el cual dio una primera mirada de la alta concentración de negatividad respecto al proceso y al organismo redactor, para después establecer una relación entre este mismo sentimiento y los *hashtags* relativos al Rechazo, finalmente hallando que los convencionales pertenecientes al sector de derecha, y cercanos al pinochetismo, fueron quienes

más reacciones negativas tuvieron durante el lapso estudiado, también descubriendo que los convencionales del Apruebo fueron en su mayoría mal evaluados por la ciudadanía. Luego, el análisis de emociones concluyó que el enojo fue la emoción predominante en el transcurso del proceso, que se tradujo también en los *hashtags* más recurrentes, los cuales también estuvieron estrechamente relacionados al Rechazo. La parte final, que trataba sobre la detección de discurso de odio, logró probar que la aprobación de la norma constitucional que establecía que Chile pasaría a ser una nación “regional, plurinacional e intercultural”, fue uno de los principales factores para que aumentara la presencia del odio en el debate constitucional, y por consecuencia empobreció la imagen de la convención constituyente, explicando de esta manera, la tendencia al alza de la negatividad entre los meses de febrero y mayo, cuyas razones no habían podido ser dilucidadas hasta ahora. Finalmente, se comprobó que los *hashtags* relacionados al discurso de odio también estuvieron relacionados al Rechazo, marcando una vez más, las tendencias que se identificaron y solidificaron con cada análisis. A continuación, para presentar todos los hallazgos desde un punto de vista más visual y menos estadístico, se confeccionaron visualizaciones de datos que le permitirán al lector entender de mejor manera los resultados obtenidos en la presente sección.

3.6. Interpretación de resultados

Finalmente, luego de haberse completado todas las etapas de la metodología KDD previamente mencionadas y obtenidos sus resultados, se construyeron diferentes visualizaciones de datos como un motor para interpretar los datos y el análisis realizado sobre ellos para llevar este estudio un paso más allá y poder llegar al público general con el fin de que todos sean capaces de comprender un poco más lo que realmente fue el debate constitucional en Chile durante el 2022.

Herramientas como tablas, gráficos de barras, de líneas, radiales, etc. fueron útiles para apreciar de una forma más visual la evolución del debate, las tendencias que se fueron instalando y los eventos que marcaron cambios en éstas, los personajes más importantes que participaron de éste, su rol dentro del mismo y el grado de influencia que tuvieron sobre los demás.

Para construir las visualizaciones previamente mencionadas se utilizó Flourish, una plataforma gratuita que ofrece más de 60 tipos diferentes de gráficos, grafos, tablas y otros tipos de visualizaciones la cual destaca por su facilidad para la elaboración de éstas debido a que sólo necesita los datos para su construcción, sin necesidad de procesamientos previos ni posteriores. Además, importar datos a esta plataforma es sencillo, gracias a que soporta la mayoría de los formatos comunes de *datasets* como hojas de cálculo de Excel o csv (*comma separated values*), siendo este último el formato con el que se trabajó.

El panorama que presenta esta investigación hizo necesario el uso de un amplio rango de elementos que permitieran ilustrar, de una manera aún más visual, los resultados y hallazgos encontrados, utilizando herramientas como nubes de palabras e incluso agregando canales visuales como el color que ayudaran también a entender el sentimiento comprendido detrás

de éstos, teniendo siempre cuidado de que las visualizaciones fuesen intuitivas y hablaran por sí mismas.

3.6.1. Distribución del sentimiento

Mediante gráficos de barras horizontales se ilustró la distribución del porcentaje de *tweets* identificados con sentimientos negativos, neutrales o positivos. El color predominante seleccionado para las visualizaciones fue el azul, con tonos más claros indicando la positividad y los tonos más oscuros indicando negatividad.

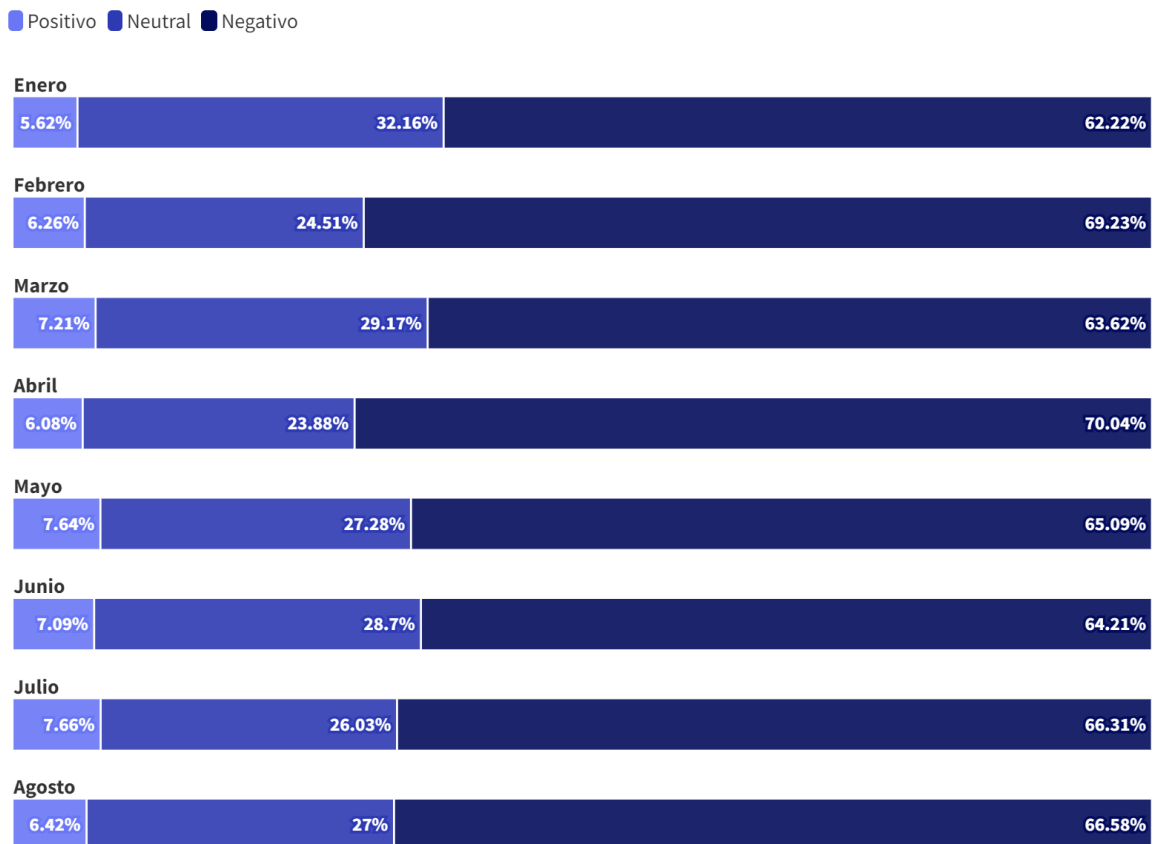


Figura 8: Distribución del sentimiento en el *dataset* sin respuestas.
Fuente: Elaboración propia.

■ Positivo ■ Neutral ■ Negativo

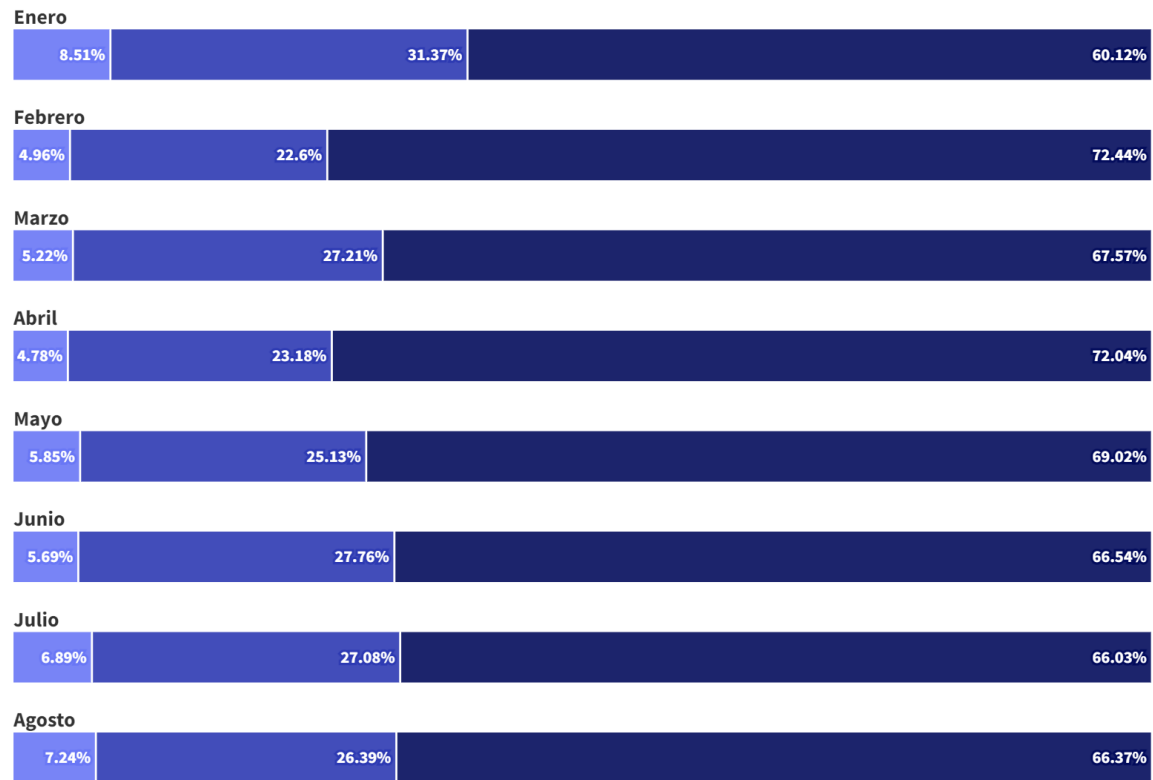


Figura 9: Distribución del sentimiento en el *dataset* con respuestas.
Fuente: Elaboración propia.

La Figura 8 muestra la distribución del porcentaje de *tweets* clasificados como positivos, neutrales y negativos, ordenados de izquierda a derecha y con los colores previamente mencionados, en el cual claramente se puede apreciar la predominancia de los *tweets* negativos, seguido de los neutrales y finalmente los positivos. Si se mira verticalmente el gráfico, es posible apreciar la evolución mes a mes de los tres sentimientos, pudiendo identificar un alza en la negatividad en febrero en desmedro de la neutralidad, que luego se recuperaría en marzo, únicamente para bajar de nuevo, y con mayor intensidad en abril. Por su parte, la Figura 9 muestra la misma distribución pero para el *dataset* con respuestas, en el cual se puede apreciar ligeramente una mayor predominancia de los *tweets* negativos, al igual que una mayor heterogeneidad en las distribuciones de los sentimientos entre enero y mayo, que presentan mayor disparidad respecto a la visualización anterior, correspondiente al *dataset* sin respuestas.

3.6.2. Evolución del sentimiento a través del tiempo

Las dos visualizaciones presentadas anteriormente hacen un buen trabajo en ilustrar la distribución del sentimiento mes a mes, sin embargo, es difícil apreciar la evolución que éstas tienen a través del tiempo. Es por este motivo que las visualizaciones que se presentarán a continuación corresponden a gráficos de líneas que emplearán la misma paleta de colores usada en las visualizaciones de la sección anterior.

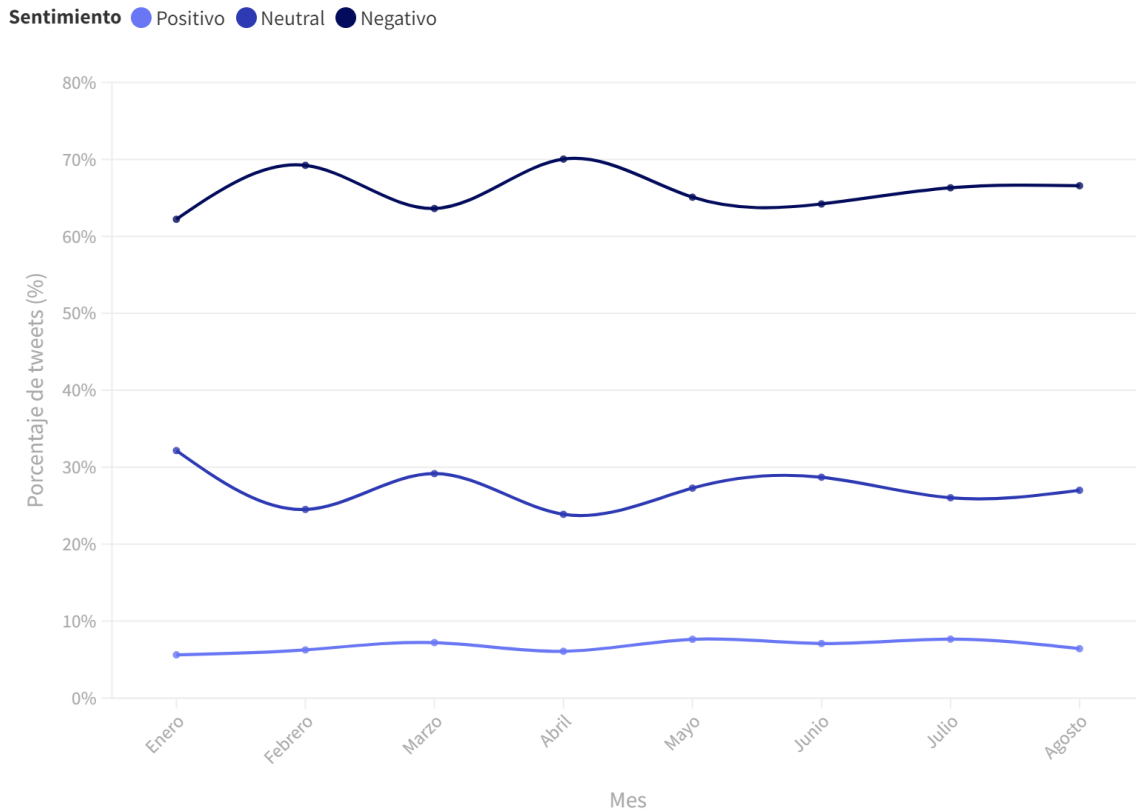


Figura 10: Evolución del sentimiento a través del tiempo en el *dataset* sin respuestas.
Fuente: Elaboración propia.

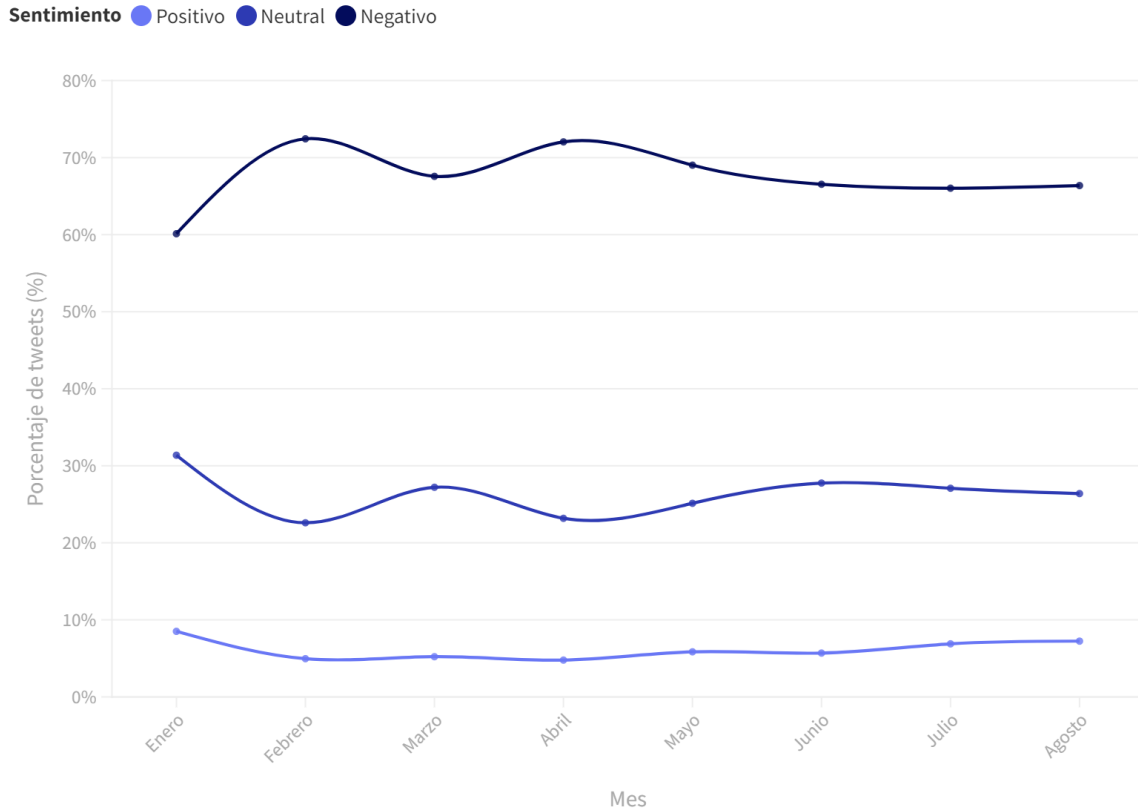


Figura 11: Evolución del sentimiento a través del tiempo en el *dataset* con respuestas.
Fuente: Elaboración propia.

Como se puede apreciar, las visualizaciones presentes en las Figuras 10 y 11 tienen el mismo propósito, pero trabajan con datos diferentes, recordando que la Figura 10 presenta la evolución a través del tiempo para el *dataset* sin respuestas, mientras que la Figura 11 lo hace para el que contiene los *tweets* originales más sus respuestas. La única diferencia identificable a simple vista es que la curva de los *tweets* negativos está ligeramente desplazada hacia arriba en el *dataset* con respuestas respecto al que no tiene, lo cual es consistente con los porcentajes mostrados en las visualizaciones de la sección anterior, en donde se señala que la cantidad de *tweets* negativos es superior en el *dataset* de mayor volumen.

También es importante mencionar que es posible apreciar, de mejor manera, la evolución del sentimiento a través del tiempo mediante la observación de la curva correspondiente a cada sentimiento en particular, sobre todo en los casos de los *tweets* negativos y los neutrales, ya que en los positivos no se observaron cambios bruscos entre meses. En el caso de los dos primeros, es posible observar que sus curvas son inversamente proporcionales, es decir, cuando sube el porcentaje de *tweets* negativos, baja la cantidad de *tweets* neutrales y viceversa, entregando un nuevo hallazgo que no había sido posible encontrar anteriormente mediante la información sin graficar. A través de la visualización de datos, es posible concluir

que las opiniones fueron fluctuando entre negativo y neutral, lo cual podría calificarse como votantes “indecisos”, término utilizado en encuestas electorales. Otro punto importante a señalar, es que las curvas de los tres sentimientos presentados tienden a normalizarse a partir del mes de junio, donde ya no se observan fluctuaciones considerables en la evolución del sentimiento.

3.6.3. Comparación de la evolución del sentimiento a través del tiempo

En la sección correspondiente al análisis de sentimiento, se mencionó la ventaja de seccionar los *datasets* trabajados por mes, con el objetivo de lograr visualizar su evolución a través del tiempo, que se realizó en las visualizaciones recientemente mostradas, pero también con el fin de poder realizar un análisis comparativo entre las curvas presentadas por ambos *datasets*. La presente sección ilustra los resultados del sentimiento identificado en ambos conjuntos de datos y cómo sus mínimas diferencias decantaron en la decisión de no seguir trabajando con ambos, sino que sólo con el *dataset* con respuestas.

La visualización que se muestra utiliza la paleta de colores basada en el azul como color primario para representar las curvas mostradas por el *dataset* con respuestas, mientras que se introducirá una nueva paleta basada en el violeta para el *dataset* sin respuestas.

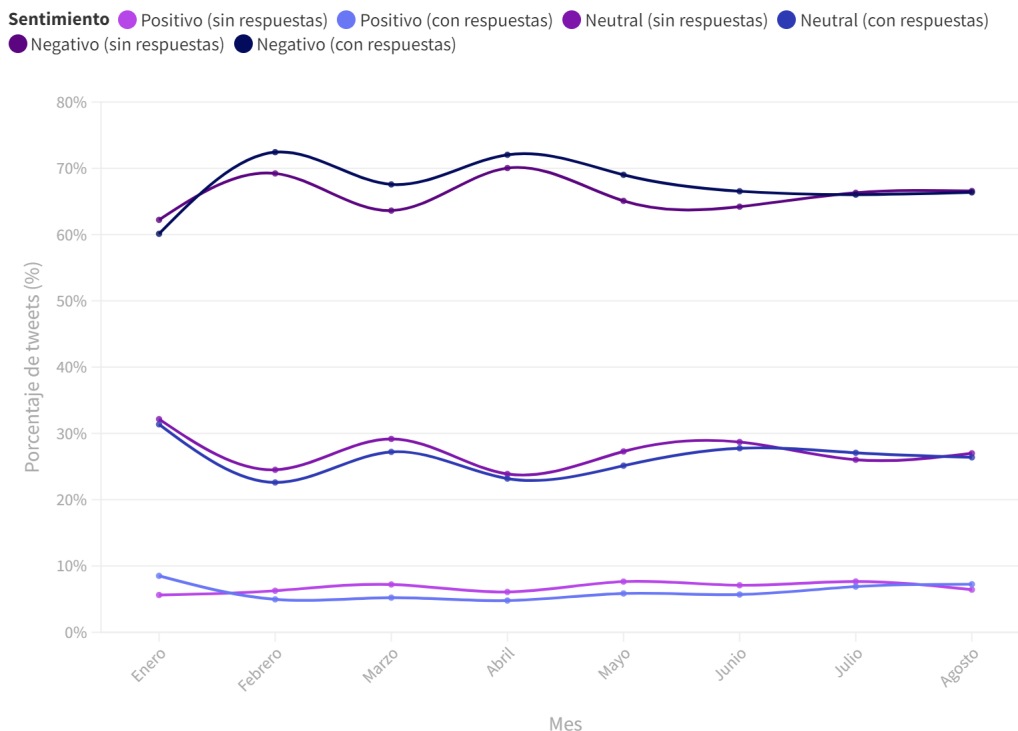


Figura 12: Comparación de la evolución del sentimiento a través del tiempo.

Fuente: Elaboración propia.

En la Figura 12 se puede apreciar claramente la diferencia entre las curvas obtenidas para los tres sentimientos diferentes, avalando la decisión tomada en la sección referente al análisis de sentimiento, en donde se desestimó el uso de los dos *datasets* durante el resto del estudio debido a que no se hallaron grandes diferencias. Como se puede ver, las curvas para el porcentaje de *tweets* positivos y neutrales son prácticamente idénticas, mientras que la negativa se ve un poco más desplazada hacia arriba en la correspondiente al *dataset* con respuestas, lo cual ya se había señalado en párrafos anteriores.

Las siguientes visualizaciones a mostrar fueron confeccionadas en base únicamente al *dataset* que contiene los *tweets* originales extraídos durante el proceso más sus respuestas, puesto que es más representativo de la percepción general de los *twitteros* respecto al proceso constitucional.

3.6.4. Positividad, neutralidad y negatividad

Durante el transcurso del análisis de sentimiento se habló acerca de métricas conocidas como positividad, neutralidad y negatividad, que corresponden a los pesos ponderados obtenidos por cada sentimiento, ofreciendo una visión más realista del parecer de los usuarios de Twitter respecto a la convención constitucional. Al igual que en las visualizaciones anteriores, también se graficó su distribución mes a mes, siguiendo sus mismas reglas en cuanto a orden, gráfico utilizado y paleta de colores. Hay que recordar que previo a estos resultados se desechó el uso del *dataset* sin respuestas, por lo que las visualizaciones a continuación fueron hechas a partir de los hallazgos obtenidos en el estudio del *dataset* que incluye las respuestas y sus *tweets* originales.

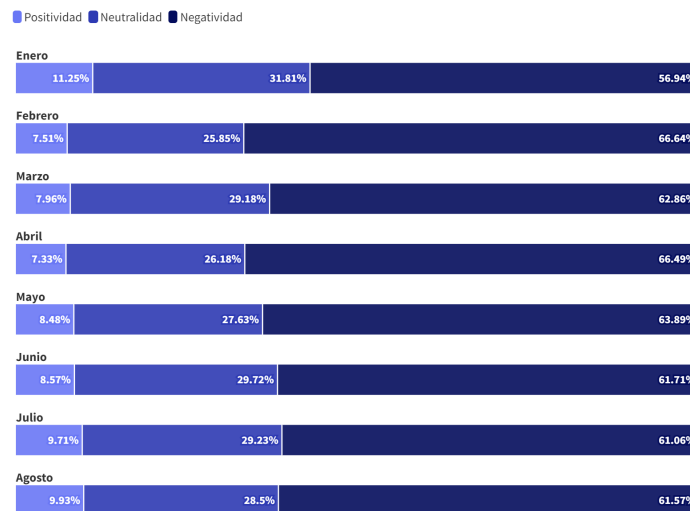


Figura 13: Distribución de la positividad, neutralidad y negatividad.
Fuente: Elaboración propia.

Como se puede ver en la Figura 13, se presenta la misma naturaleza heterogénea de la distribución entre los meses de enero y mayo presentada en la Figura 9, pero con una positividad bastante mayor a la presentada en ésta, y con la negatividad jamás superando el 70%. Sin embargo, también es interesante ver cómo evolucionaron la positividad, neutralidad y negatividad a través del tiempo en el lapso estudiado.

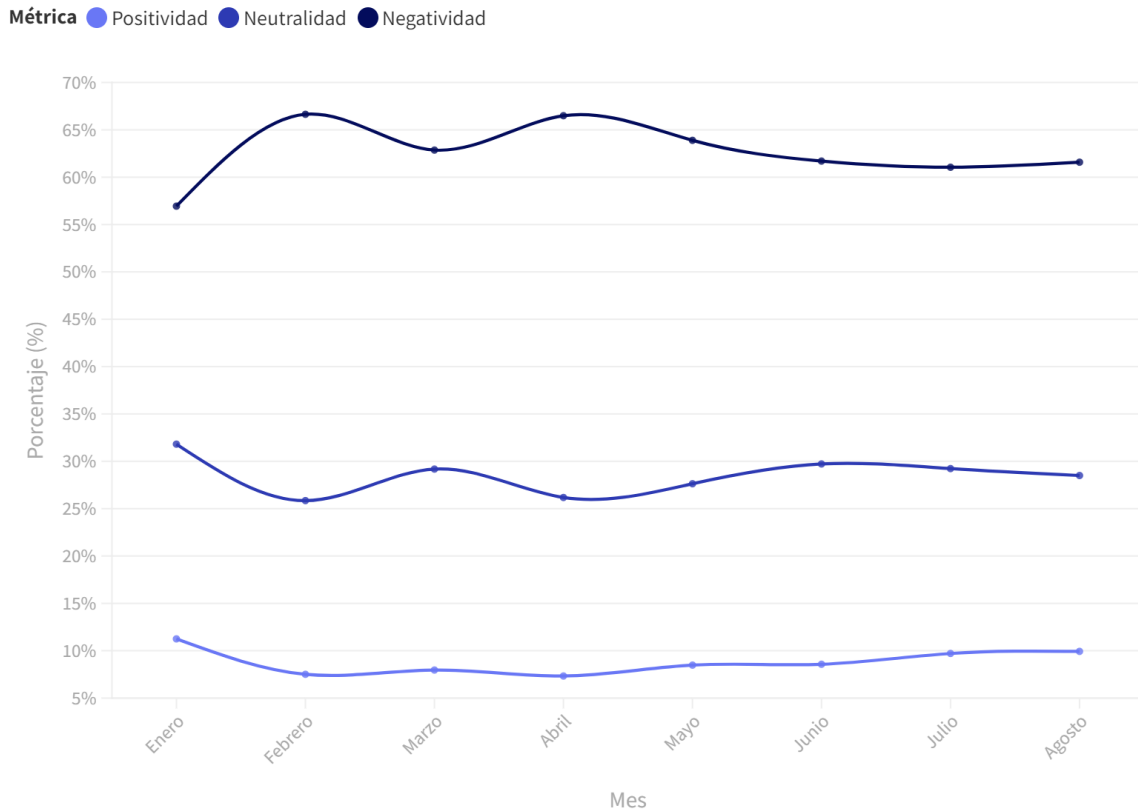


Figura 14: Evolución de la positividad, neutralidad y negatividad a través del tiempo.
Fuente: Elaboración propia.

La Figura 14 muestra claramente lo discutido anteriormente, lo relativo a las curvas de negatividad y positividad, que se ven afectadas y distintas respecto a los gráficos anteriores, viendo como la curva de la positividad se encuentra cercana al 10% y la de la negatividad firme en el rango 60%-70%, manteniendo la tendencia que se había visto anteriormente. Es importante notar que la positividad, que anteriormente se había mostrado como una curva constante, casi una línea recta, ahora presenta una pequeña curva descendiente entre los meses de febrero y abril, que después comienza a recuperarse en el mes de mayo, similar a la tendencia encontrada en la sección de detección de discurso de odio y lo relacionado a la norma de Chile plurinacional e intercultural.

claramente que los *hashtags* #Rechazo y #ConvencionConstitucional fueron los que más repitieron en este lapso, pudiendo visualizar también nuevos *hashtags* que no se habían visto en análisis anteriores, como *hashtags* más asiduos al Apruebo, teniendo como ejemplo a #LaConvencionSeDefiende; *hashtags* sobre sucesos momentáneos, como #QuintoRetiro, referenciando el quinto retiro de las AFP, que se discutió durante dicha época en la Cámara de Diputados, o #LonconYBassaMV, *hashtag* utilizado para la aparición de Elisa Loncón y Jaime Bassa, presidenta y vicepresidente de la convención constitucional en ese entonces, en el programa Mentiras Verdaderas, emitido por el canal La Red. Y finalmente, más *hashtags* golpeando a la izquierda como sector político y al actual gobierno de Gabriel Boric, en la forma de #BoricDictador y #GobiernoDeInutiles.

No obstante, es más importante ver cómo estos *hashtags* se dividieron por sentimiento, para visualizar de mejor manera cómo se distribuyeron por cada uno de éstos en el período y dónde fueron más utilizados finalmente, lo cual se muestra en la nube de la Figura 16, que esta vez utilizará la paleta de colores con el azul como color primario, para hacer la distinción por sentimiento, pero utilizando esta vez el color rojo para representar a los *tweets* negativos, ya que los *hashtags* no tan recurrentes saldrán con un tamaño menor en la nube, lo que podría generar confusión en la identificación de los sentimientos. Se puede ver, con claridad, la separación por sentimiento de los *hashtags*, viendo una nube mayoritariamente roja, marcando la predominancia de la negatividad en los *hashtags* más utilizados, como #Rechazo y todas sus variaciones, que ocurre de igual manera, pero con menor magnitud en la positividad para el #Apruebo y sus variantes. En cuanto a los *hashtags* neutrales, destacan #ConvencionConstitucional y sus variantes, junto con nombres de programas principalmente radiales, como #CooperativaContigo y #ElMostradorEnLaClave, cuya objetividad como medios de comunicación se condice con una posición neutra respecto del trabajo de la convención constitucional.

ANÁLISIS DE LOS HECHOS Y TENDENCIAS QUE DETERMINARON EL RESULTADO DEL PLEBISCITO CONSTITUCIONAL DE 2022, MEDIANTE PERIODISMO DE DATOS

All

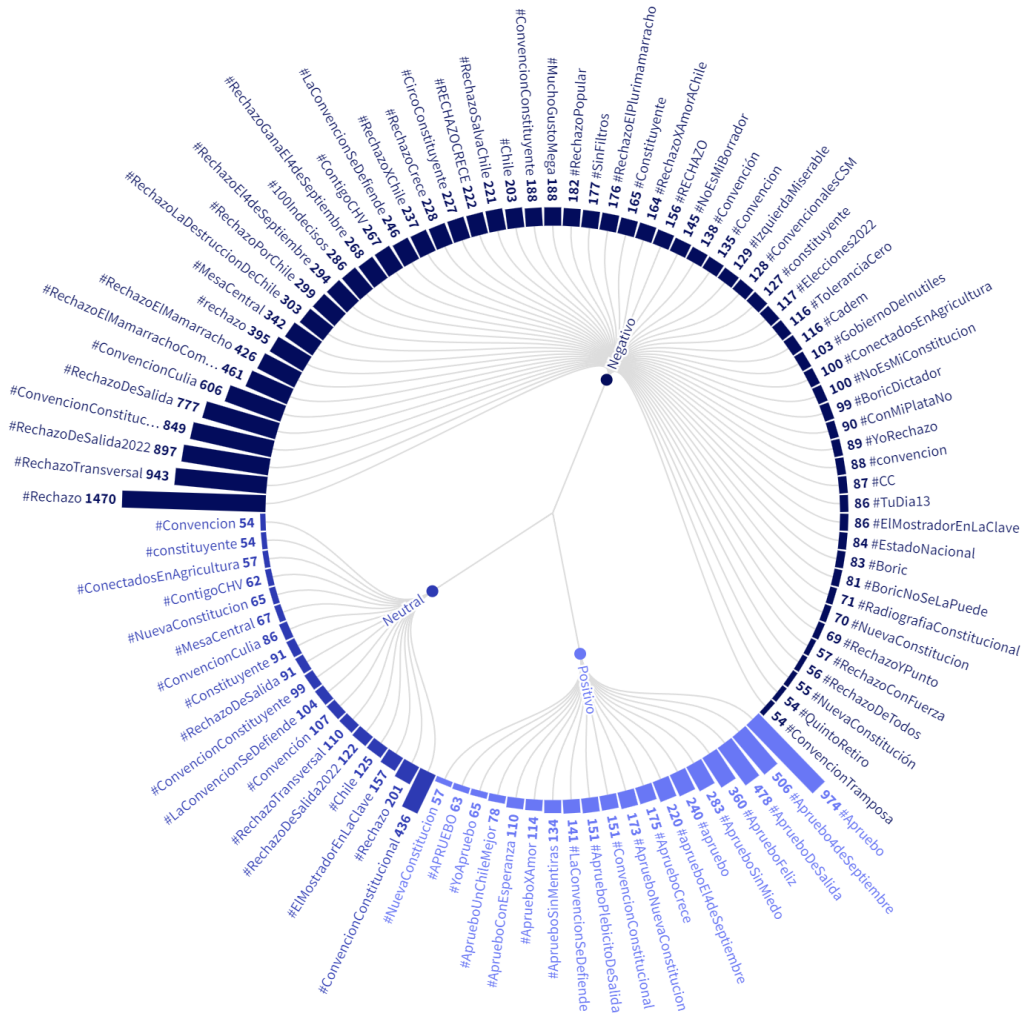


Figura 17: Hashtags más recurrentes y sus frecuencias durante los meses de enero y septiembre de 2022, clasificados por sentimiento.

Fuente: Elaboración propia.

La visualización presentada en la Figura 17 permite ver de una manera mucho más clara la distribución de los hashtags por sentimiento, mostrando la predominancia de los hashtags “negativos”, que toman más de la mitad de la circunferencia del círculo, dejando el resto a los hashtags neutrales y positivos, los que parecen distribuirse de manera similar en cuanto a cantidad individual o perímetro, pero no en volumen, dado por la altura de cada barra que se desprende del borde del círculo, siendo las pertenecientes al lado positivo mucho más

altas que las neutrales, concluyendo que, en cuanto a *hashtags*, el sentimiento positivo se sobrepuso al neutral, pero sigue estando muy lejos del negativo, que lo triplica en tamaño.

3.6.6. Los actores del debate y su rol e influencia en este

La última sección del análisis de sentimiento se enfocó fuertemente en analizar el rol de los actores del proceso, sean constituyentes, personajes públicos o medios de comunicación masivos, con el fin de establecer quiénes fueron los más importantes en cuanto a influencia sobre los demás. Sin embargo, la estadística sola no es capaz de entregar el panorama completo de lo que en realidad significa develar el rol y la influencia de los actores del debate sobre las demás personas, que generalmente forman su opinión en base a las declaraciones de éstos. Mediante visualizaciones de datos, se intentó ilustrar dicho panorama mediante gráficos que manifiesten de mejor manera lo que se quiso lograr en este apartado.

La Figura 18 muestra un gráfico de barras ordenado de mayor a menor en negatividad para los 10 constituyentes más mencionados en Twitter durante el proceso, utilizando la paleta de colores habitual basada en el color azul. Este gráfico presenta claramente la disparidad en sentimiento entre los convencionales de izquierda y de derecha, ubicándose tres de cinco constituyentes de derecha (Zúñiga, Fontaine y Marinovic) en la parte superior, mientras que los cuatro convencionales con menor cantidad de menciones negativas pertenecen a la izquierda, lo cual se condice con el hecho de que la derecha chilena, como sector, se manifestó en contra del proceso incluso antes de haberse iniciado, mientras que la izquierda fue la principal impulsora de éste. Sin embargo, es importante reconocer que todos los constituyentes presentes en el gráfico presentan cantidades muy altas de menciones negativas, independiente del sector político, que a su vez es consistente con la mala percepción en general de los usuarios de Twitter respecto a la convención constitucional.

ANÁLISIS DE LOS HECHOS Y TENDENCIAS QUE DETERMINARON EL RESULTADO DEL PLEBISCITO
CONSTITUCIONAL DE 2022, MEDIANTE PERIODISMO DE DATOS

■ Negativo ■ Neutral ■ Positivo

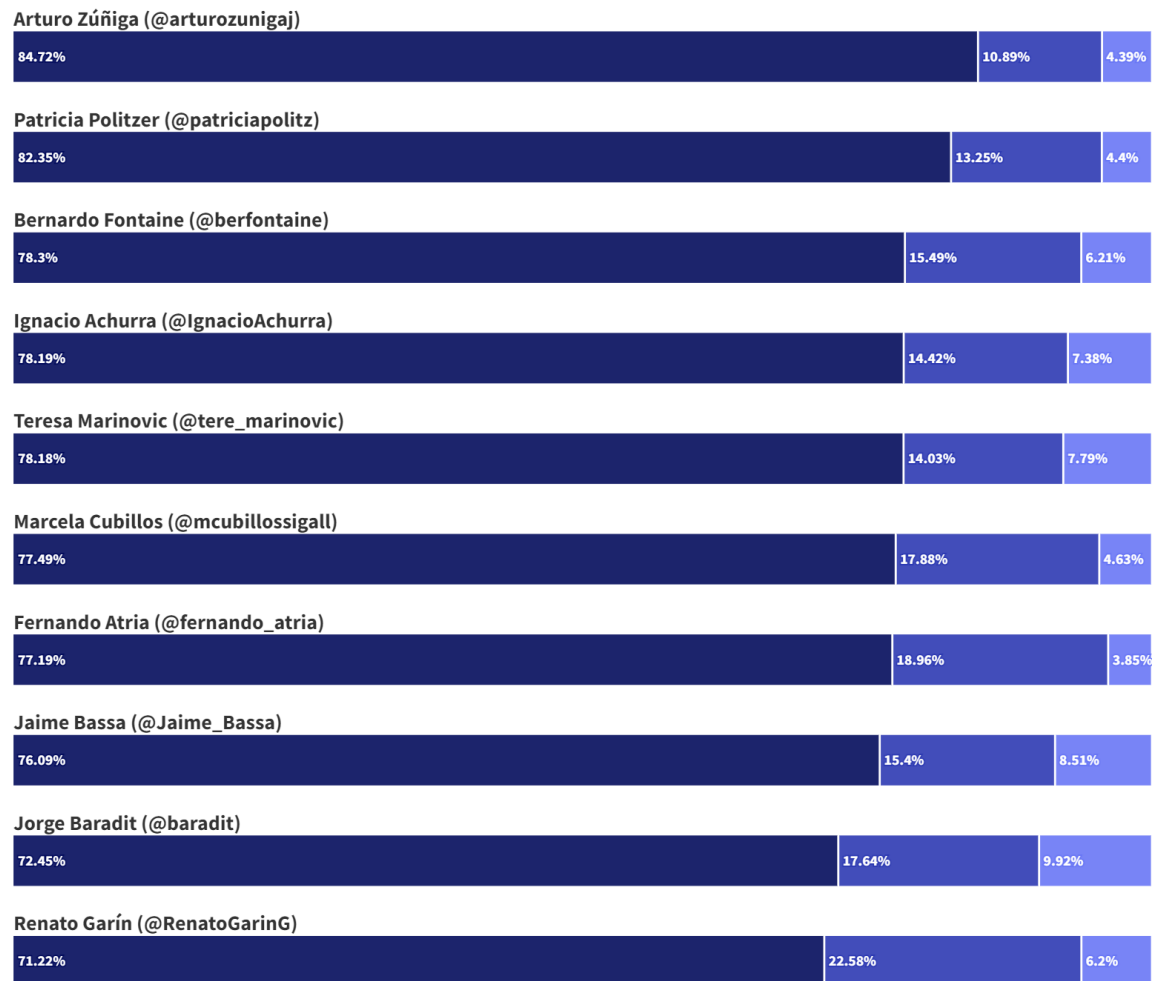


Figura 18: Top 10 constituyentes más mencionados en Twitter, clasificados por el sentimiento vertido en sus menciones.

Fuente: Elaboración propia.

Tal como en la sección del documento que se dedicó al estudio de los actores del debate, los constituyentes no fueron los únicos personajes que se llevaron la atención del público, sino que también hubo una importante influencia por parte de personajes relevantes en la política chilena, que a través de sus opiniones y acciones, también generaron reacciones en Twitter, las cuales al ser clasificadas por sentimiento se presentan en la Figura 19. Al igual que en el gráfico de la Figura 18, los personajes públicos de derecha se ubican en los porcentajes más altos de negatividad, de hecho los tres primeros, Melnick, José Antonio Kast y Felipe Kast pertenecen a dicho sector político. Sin embargo, aquí se da la particularidad de que personajes más ligados a la izquierda, como el presidente Gabriel Boric y Mr. Wolf se ubican en la mitad del *ranking* y no en los últimos lugares como se podría esperar, o al menos, como

ocurrió en la Figura 18. Ahora bien, la explicación de esto puede encontrarse en que el debate constitucional se vio mezclado con muchas críticas al gobierno de Gabriel Boric, razón por la cual los niveles de negatividad de sus menciones puede mostrarse como está. En el caso de Mr. Wolf, el hecho de ser un personaje anónimo que se vio constante amenazado de gravedad por seguidores de José Antonio Kast, también puede haber influido en sus resultados.

■ Negativo ■ Neutral ■ Positivo

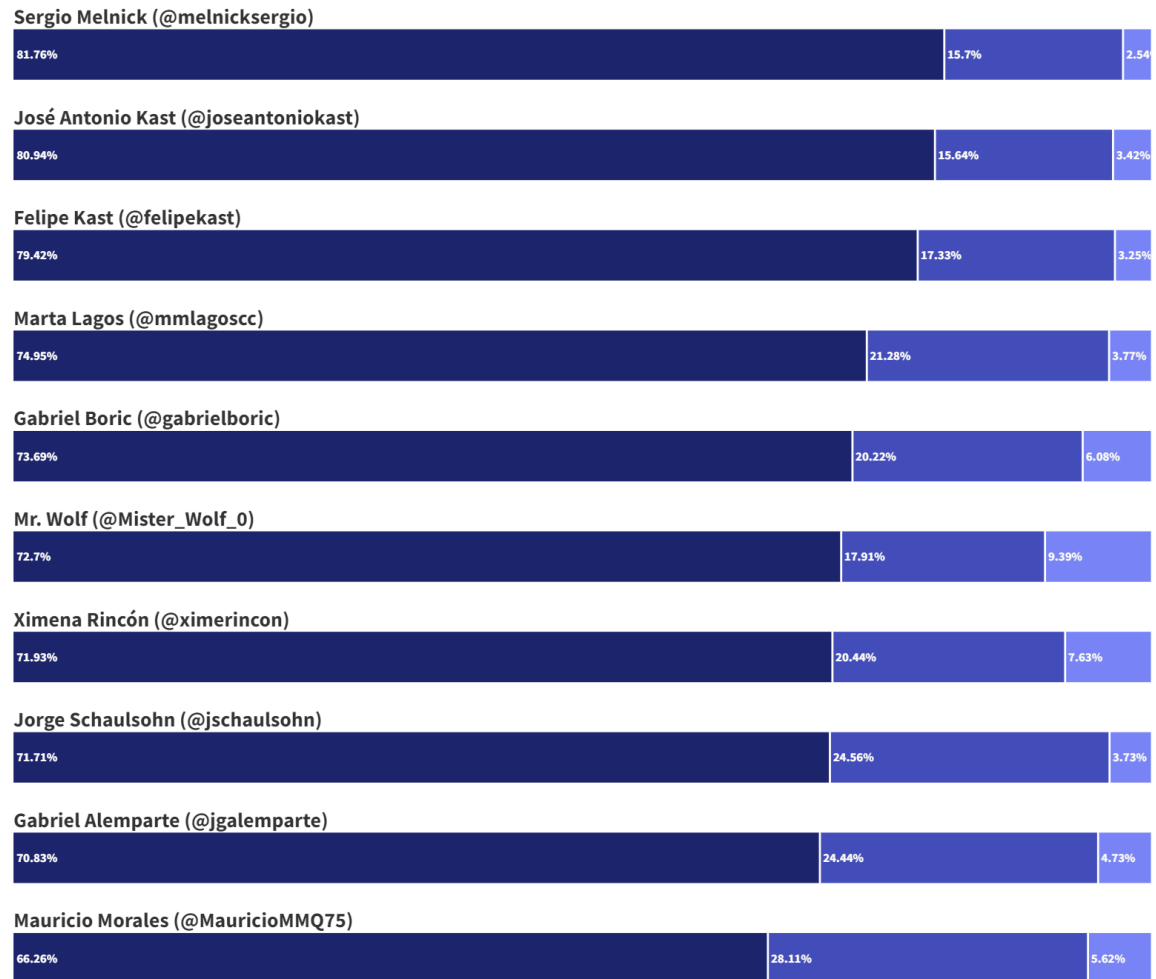


Figura 19: Top 10 personajes públicos no constituyentes más mencionados en Twitter, clasificados por el sentimiento vertido en sus menciones.

Fuente: Elaboración propia.

Al igual que en el gráfico anterior, los personajes públicos de derecha se ubican en los porcentajes más altos de negatividad, de hecho, los 3 primeros, Melnick, José Antonio Kast y Felipe Kast pertenecen a dicho sector político. Sin embargo, aquí se da la particularidad de que personajes más ligados a la izquierda, como el presidente Gabriel Boric y Mr. Wolf se ubi-

can en la mitad del *ranking* y no en los últimos lugares como se podría esperar, o al menos, como ocurrió en la visualización 18. Ahora bien, la explicación de esto puede encontrarse en que el debate constitucional se vio mezclado con muchas críticas al gobierno de Gabriel Boric, razón por la cual los niveles de negatividad de sus menciones puede mostrarse como está. En el caso de Mr. Wolf, el hecho de ser un personaje anónimo que se vio constante amenazado de gravedad por seguidores de José Antonio Kast también puede haber influido en sus resultados.

A continuación se presentarán gráficos de dispersión que posicionan a los constituyentes y a los personajes públicos por separado de acuerdo a su posición política y el sentimiento general evocado, ajustando la escala para que los personajes con mayor porcentaje de positividad sean considerados con una percepción positiva respecto al resto, con el fin de hacerla más representativa.

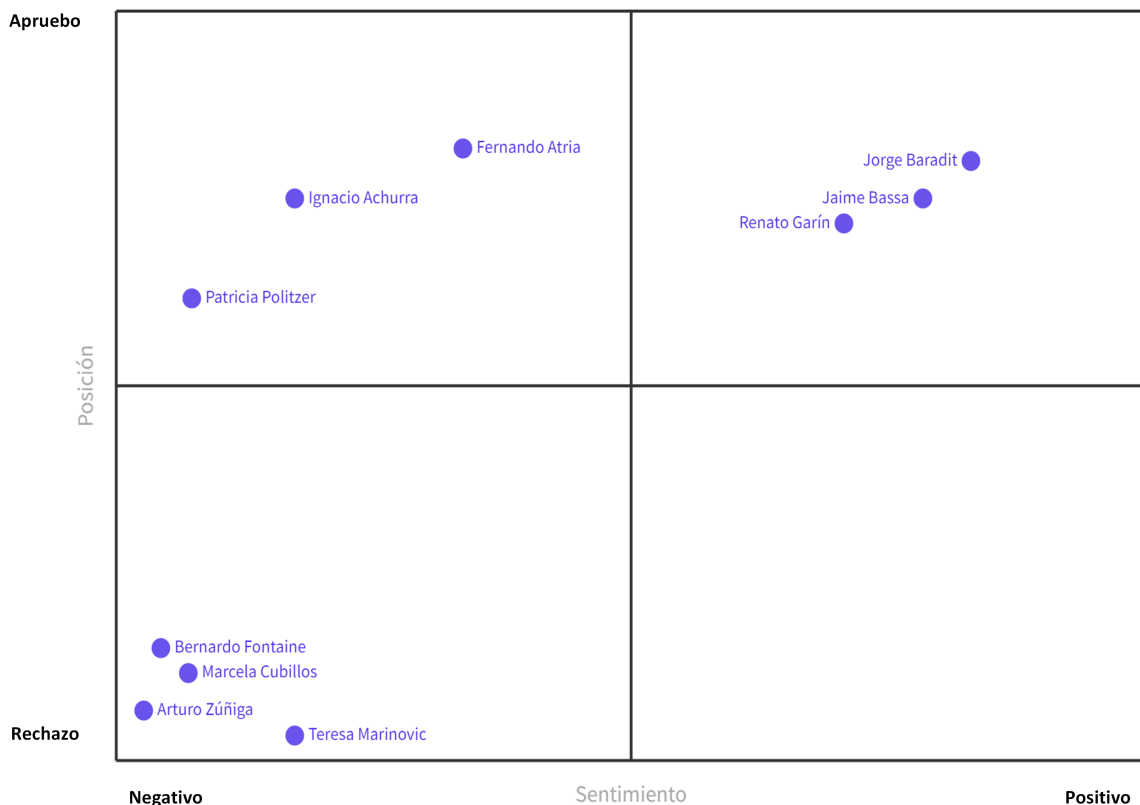


Figura 20: Top 10 constituyentes más mencionados en Twitter, posicionados de acuerdo al sentimiento vertido en sus menciones y su posición política.

Fuente: Elaboración propia.

La Figura 20 muestra una dispersión en dos ejes para ubicar a los 10 consituyentes más mencionados en Twitter en cuatro cuadrantes diferentes. El eje horizontal representa el senti-

miento en general inspirado por cada constituyente, ubicándose el negativo a la izquierda y el positivo a la derecha; por su parte, el eje vertical indica la posición política de cada candidato (Apruebo o Rechazo), métrica obtenida mediante el estudio de los tweets positivos y negativos para cada uno, con el fin de establecer una posición aproximada, tomando en cuenta también sus partidos de militancia, en caso de existir. En este eje, la parte superior indica una posición asidua al Apruebo, mientras que la región inferior señala cercanía con el Rechazo. De esta manera, fue posible ubicar a estos 10 constituyentes, cuya visualización permite sacar varias conclusiones respecto al tema.

En primer lugar, se puede ver que los 4 constituyentes de derecha, ubicados en el cuadrante inferior izquierdo, se encuentran bastante cerca unos de otros y no se acercan a otros cuadrantes, dando a entender que actuaron en bloque y fueron consistentes con sus posturas y sentimientos evocados a lo largo del proceso. En la otra vereda, los constituyentes más relacionados a la izquierda se encuentran en la parte superior, es decir, asiduos al Apruebo pero dispersos entre sentimiento, lo cual se debe a los altos niveles de negatividad presentados por todos los constituyentes que se ubican en la región superior izquierda, que no lograron el mismo efecto de sus pares que se ubicaron en la región contraria. Finalmente, es importante mencionar también que los constituyentes ubicados en el cuadrante superior derecho, si bien lograron una mejor percepción positiva que los demás, no tienen el mismo nivel de consistencia que sus pares de derecha, ubicados casi en la esquina inferior izquierda del gráfico, ya que se ubican casi al centro del cuadrante, es decir, no lograron influir a la sociedad de manera lo suficientemente positiva hacia la convención constituyente, mientras que sus equivalentes de derecha lograron inclinar la balanza hacia su lado, resultando finalmente en el triunfo del Rechazo.

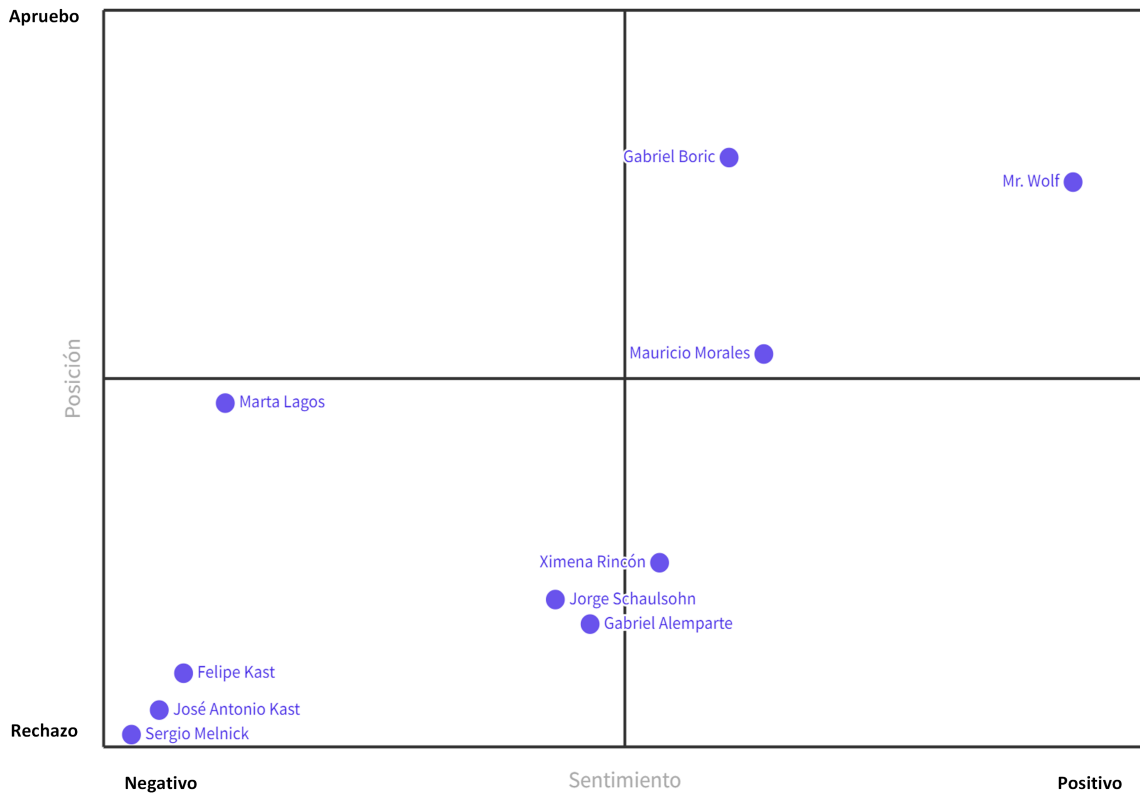


Figura 21: Top 10 personajes públicos no constituyentes más mencionados en Twitter, posicionados de acuerdo al sentimiento vertido en sus menciones y su posición política.

Fuente: Elaboración propia.

En la Figura 21 se presenta el posicionamiento para los personajes públicos que no fueron actores directos del proceso, es decir, no fueron constituyentes, en donde se puede ver que nuevamente los personajes más alineados a la derecha se mantuvieron en bloque, concentrando a José Antonio Kast, Felipe Kast y Sergio Melnick en la esquina inferior izquierda, mientras que los personajes más cercanos a la izquierda se dispersaron aún más, teniendo a un Gabriel Boric muy cercano al eje central, debido a la gran cantidad de reacciones negativas que éste generó, mientras que Mr. Wolf, el más positivo de los 10, ocupa el lugar más lejano de su extremo. A diferencia de la Figura 20, aquí aparece el llamado “centro político”, personajes con tendencias políticas poco definidas, pero que se mostraron adeptos al Rechazo, en los casos de Rincón, Schaulsohn y Alemparte, quienes se encuentran en la región inferior del gráfico, aunque muy por sobre los convencionales de derecha, que a su vez se ubicaron casi en el centro del eje horizontal referente al sentimiento, debido a las reacciones divididas que estos personajes generaron, lo cual puede deberse al giro en sus posturas que se mencionó en capítulos anteriores, provocando una mezcla en los sentimientos que despertaron, recordando que las Figuras 20 y 21 se basan en el período completo estudiado, sin hacer distinción por meses, a diferencia de la mayoría de las visualizaciones aquí presentes.

3.6.7. Distribución y evolución de las emociones

Luego de concluir el análisis de sentimiento, fue momento de analizar las emociones que se hicieron presentes en el debate, donde el enojo se alzó como la emoción predominante. Sin embargo, para mayor información de cómo se distribuyeron las emociones y su evolución a través del tiempo, se confeccionaron visualizaciones de datos que ayudan a dicho propósito. En este caso, la paleta de colores cambió, dado que las emociones aquí no se encuentran necesariamente relacionadas directamente entre ellas, así como tampoco son un grado menor o mayor la una de la otra. El enojo será representado con un color rojo, dado que las emociones fuertes y negativas generalmente son asociadas a dicho color; para la alegría, será utilizado el color azul en una tonalidad clara para ser consistentes con la positividad presentada en visualizaciones anteriores. Por último, la tristeza será ilustrada mediante el gris, un color más apagado.

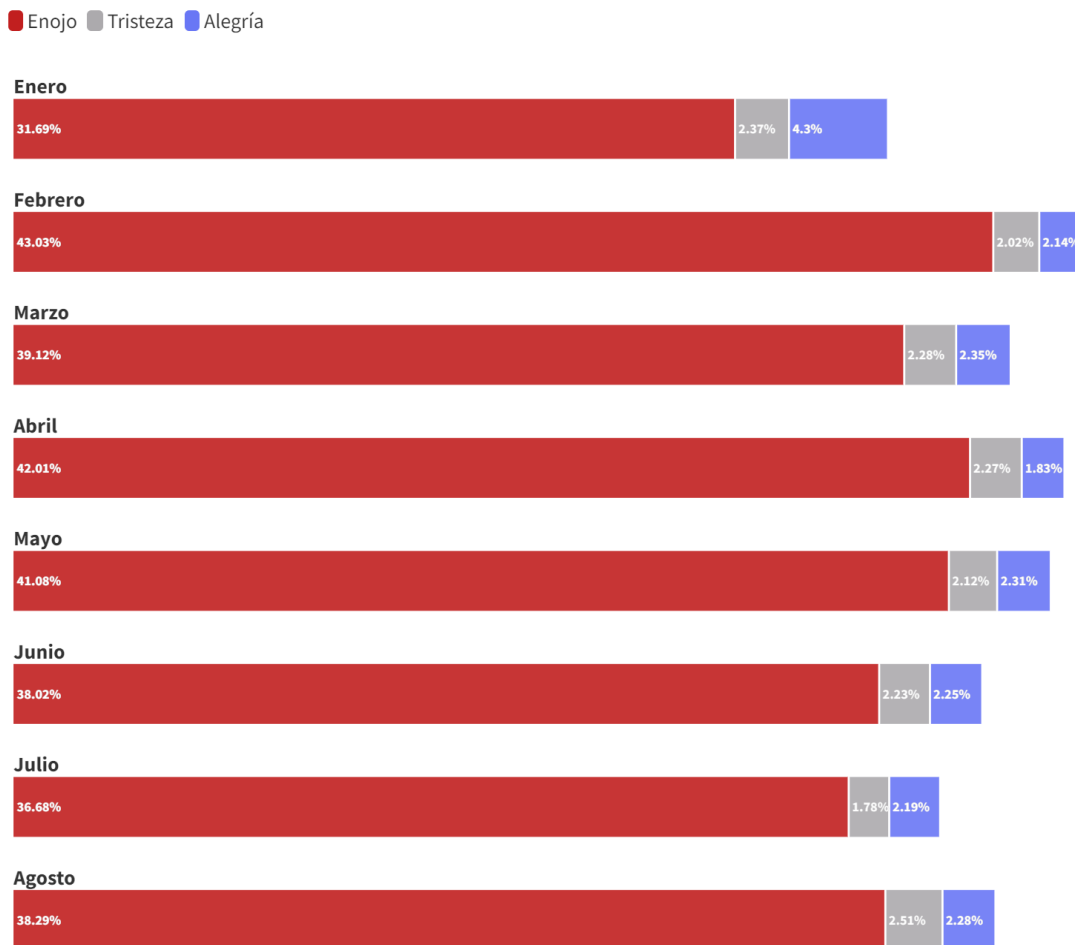


Figura 22: Distribución de las emociones.
Fuente: Elaboración propia.

En la Figura 22, se puede ver la predominancia del enojo por sobre la tristeza y la alegría, las otras dos emociones identificadas. Es importante destacar que las barras, a diferencia de los gráficos ubicados en las otras secciones, tienen diferentes largos, lo cual sucede debido a que no todos los tweets tienen una emoción predominante identificable, por ende, el porcentaje mostrado por cada mes corresponde a la fracción clasificada con determinada emoción por sobre el total de los tweets y no sobre el total de tweets con emociones identificadas.

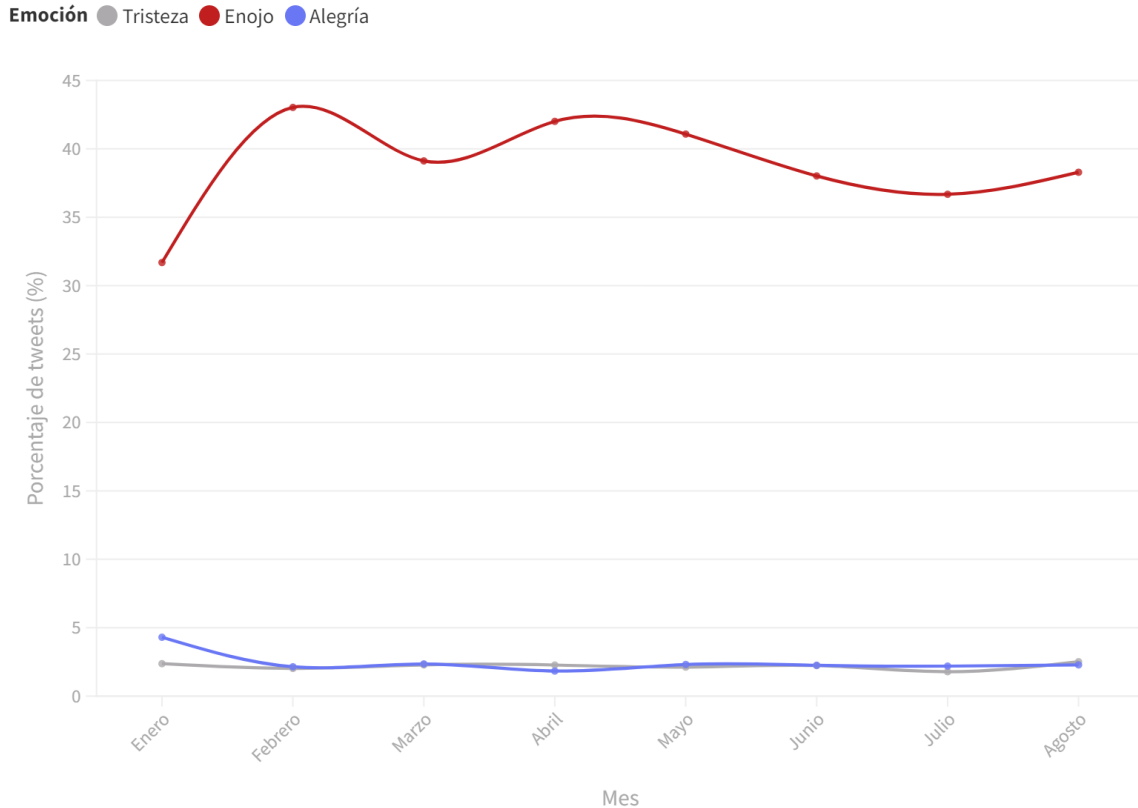


Figura 23: Evolución de las emociones a través del tiempo.
Fuente: Elaboración propia.

Durante el transcurso del análisis de emociones, se identificó una importante tendencia al alza en el enojo entre los meses de febrero y mayo, la cual en un inicio no tiene razón aparente, aunque luego en la sección de discurso de odio se descubriría que fue debido a la aprobación de la norma que declararía a Chile como una nación regional, plurinacional e intercultural. En la Figura 23, la tendencia se manifiesta en forma de una curva que tiene una forma bastante parecida a las mostradas por el porcentaje de tweets negativos y la negatividad en las Figuras 11 y 14, lo cual da a entender que la tendencia estuvo presente desde mucho antes y sólo fue en esta etapa donde finalmente se identificó. Sumando esto a la relación de proporcionalidad inversa entre la negatividad y la neutralidad hallada en la Figura 14, puede concluirse que la aprobación de dicha norma tuvo un impacto mucho más grande

del que se dimensionó anteriormente.

3.6.8. *Hashtags* y análisis de emociones

De la misma manera que se hizo en el análisis de sentimientos, los *hashtags* también fueron parte importante del estudio de las emociones, estableciendo una correlación entre los *hashtags* negativos y los que provienen de *tweets* cuya emoción predominante es el enojo. Mediante una nube de palabras en la Figura 24 y un gráfico radial en la Figura 25, se presentan los *hashtags* más recurrentes durante el período analizado, clasificados por emoción. Cabe señalar que para estas visualizaciones se mantuvieron los colores definidos para el enojo, la tristeza y la alegría.

Varios de los *hashtags* que ya habían aparecido en la Figura 16 clasificados como negativos vuelven a aparecer aquí etiquetados bajo la emoción del enojo, generalmente mediante los *hashtags* relativos al Rechazo, mientras que los *hashtags* relativos al Apruebo aparecen mayoritariamente relacionados a la alegría. Sin embargo, existe una cantidad no menor de *hashtags* relativos al Apruebo que en esta nube de palabras están calificados bajo la emoción del enojo, lo cual puede deberse a dos causas: un error en el modelo de clasificación de emociones, o *tweets* que defendieron a la convención constitucional y su trabajo con enojo. En cualquiera de los dos casos, no fue estudiado en este documento debido a que la cantidad de *tweets* identificados con estas características fue demasiado baja como para ser tomada en consideración.

3.6.9. Evolución de la presencia de discurso de odio y el impacto de la plurinacionalidad e interculturalidad

El análisis del *dataset* construido para esta investigación terminó con la detección de discurso de odio, detectando niveles consistentes con la negatividad y la presencia de enojo encontradas en análisis previos de sentimiento y emoción. Esta sección se encarga de mostrar la evolución de la presencia de discurso de odio a lo largo del tiempo, junto con mostrar el efecto de la aprobación de la norma referente a Chile plurinacional e intercultural en la intensidad de éste. Los gráficos que se mostrarán a continuación utilizaron una paleta basada en el rojo como color primario, para representar el sentimiento de odio.

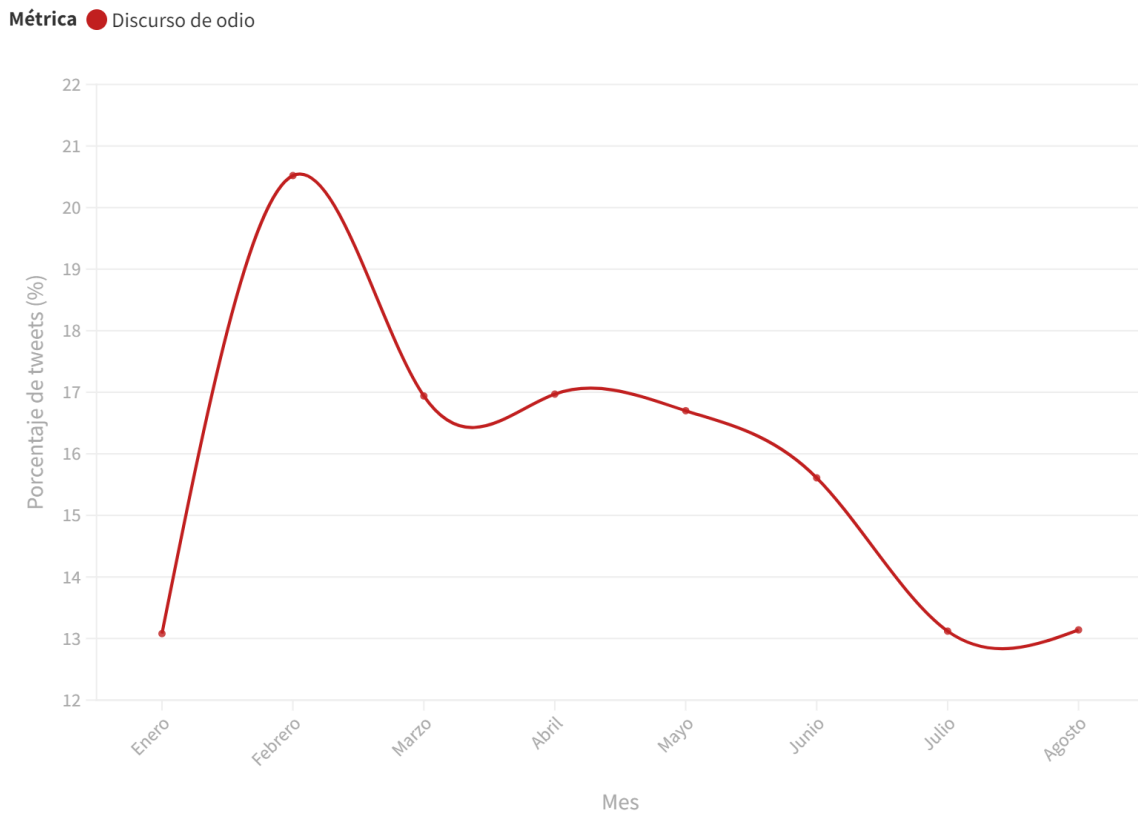


Figura 26: Evolución de la presencia de discurso de odio a través del tiempo.
Fuente: Elaboración propia.

En la Figura 26 puede apreciarse claramente la curva que se ha venido mencionando a lo largo de esta sección referente a las visualizaciones de datos, con la particularidad de ser altamente pronunciada, justificando la decisión de estudiar específicamente las razones que se encuentran detrás de esta tendencia. Como ya fue estudiado anteriormente, la razón tuvo que ver con el carácter plurinacional e intercultural que la convención constitucional quiso darle a Chile, cuyas repercusiones también llegaron a Twitter, mostrándose en la Figura 27.

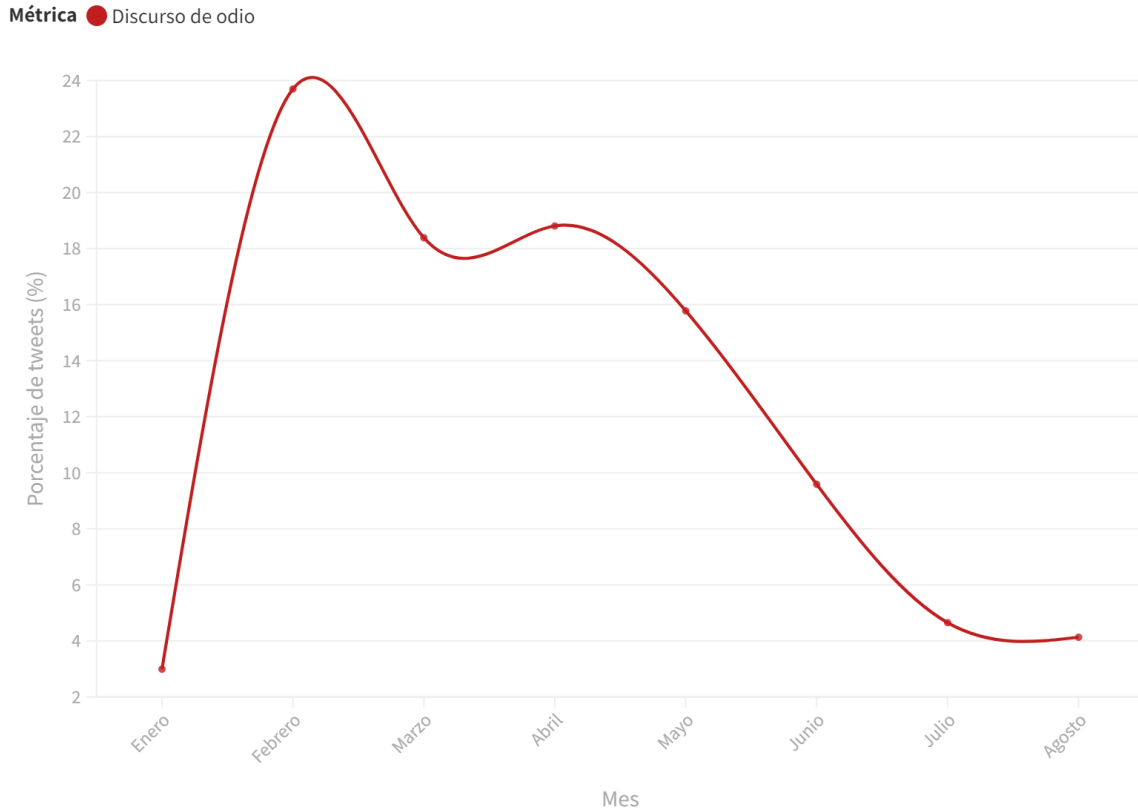


Figura 27: Evolución de la presencia de discurso de odio en *tweets* relacionados a la pluri-nacionalidad e interculturalidad a través del tiempo.

Fuente: Elaboración propia.

Dicho gráfico muestra una curva de igual forma que las ya estudiadas, coincidiendo con la presencia del discurso de odio, la presencia del enojo y la negatividad durante los meses de febrero y mayo. Para reconfirmar la hipótesis planteada, se construyó un gráfico comparativo que enfrenta la curva real de discurso de odio versus la curva sin contar los *tweets* que hablan de la norma en los que fue detectado discurso de odio.

Como se puede apreciar en la Figura 28, las curvas si bien tienen formas parecidas, la curva de color naranja, que representa a la evolución de la presencia de discurso de odio sin considerar aquellos *tweets* que hablan sobre la norma de plurinacionalidad e interculturalidad, tiene un crecimiento considerablemente menor, ilustrando de una forma más gráfica el hallazgo que se había hecho anteriormente: la aprobación de dicha norma fue una de las causas principales del deterioro de la percepción de los usuarios de Twitter sobre el trabajo de la convención constitucional.

Métrica ● Con plurinacionalidad e interculturalidad ● Sin plurinacionalidad e interculturalidad

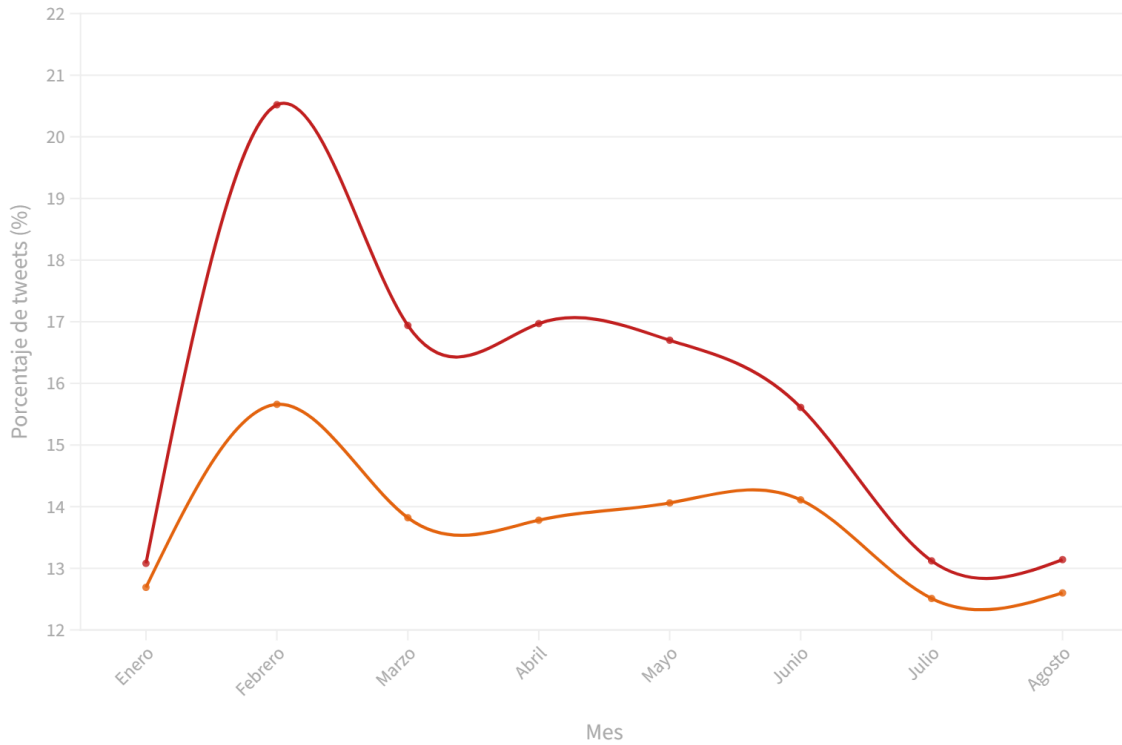


Figura 28: Comparación de la presencia de discurso de odio a través del tiempo entre la presencia y ausencia de la plurinacionalidad e interculturalidad.

Fuente: Elaboración propia.

De esta manera, finaliza esta sección, que mediante la visualización de datos logró dar una nueva mirada a lo que fue el proceso constitucional, mediante gráficos y nubes de palabras cuyo objetivo fue mostrar los aspectos más profundos tocados por esta investigación, presentándolos de una manera más intuitiva de ver, facilitándole al lector el entendimiento de los resultados de este estudio.

Así, a través de la metodología KDD fue que se construyó la solución que aporta mayor información sobre el debate constitucional en Chile durante el año 2022 a partir de las opiniones de los votantes recopiladas en Twitter y mediante el uso de herramientas como el *web scraping*, el análisis de sentimiento y la visualización de datos.

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

El presente capítulo trata acerca de los resultados obtenidos por la solución propuesta en el capítulo anterior, abarcando todo el proceso que esto conlleva, para luego pasar a validar la solución mediante la presentación de las visualizaciones de datos elaboradas al final del proceso, con el objetivo de determinar si efectivamente los resultados obtenidos amplían el panorama que el público tiene sobre el debate constitucional en Chile del 2022.

4.1. Metodología de validación

Para validar la solución entregada, se sometieron a prueba las visualizaciones de datos confeccionadas, con el fin de determinar si es que los gráficos y nubes de palabras construidos cumplen el propósito deseado. Así, se diseñó una metodología de tres etapas, cuya primera etapa consiste de un cuestionario de entrada en el cual los participantes respondieron preguntas en base a su conocimiento del proceso constitucional en Chile durante el 2022. En la segunda parte, el grupo de personas evaluó las visualizaciones construidas en base a un cuestionario preconstruido, asignando una nota o puntaje del 1 al 7, siendo el 1 en completo desacuerdo y el 7 en completo acuerdo. Y, finalmente, contestar un cuestionario de salida, el cual dimensionó el impacto de las visualizaciones de datos en el conocimiento y entendimiento del tema en los usuarios.

4.1.1. Grupo de validación

Las personas seleccionadas para la validación de la propuesta consistió en un grupo de 31 personas, de las cuales 18 fueron hombres y 13 mujeres, con un rango etario comprendido entre los 19 y los 53 años, más específicamente entre 21 y 53 años para los hombres, y 19 a 49 años para las mujeres.

4.1.2. Cuestionario de entrada

Todos los participantes del experimento debieron responder un breve cuestionario de tres preguntas con el objetivo de establecer el conocimiento inicial de cada uno de ellos respecto al proceso constitucional del 2022. Las preguntas que componen dicho cuestionario son las siguientes:

- Del 1 al 7, indica tu nivel de conocimiento respecto del proceso constitucional del 2022.

- Del 1 al 7, ¿qué tan negativa crees que era la percepción de la gente respecto a la Convención Constituyente?
- Del 1 al 7, ¿qué tan positiva crees que era la percepción de la gente respecto a la Convención Constituyente?
- A tu juicio, nombra cuáles crees que fueron las tres principales causas del triunfo del Rechazo en el plebiscito constitucional del 2022.
- ¿Crees que Twitter es una herramienta confiable para predecir resultados políticos?

4.1.3. Cuestionario de evaluación

Luego de contestar el cuestionario de entrada, a los participantes se les mostraron 6 visualizaciones de datos, correspondientes a las Figuras 9, 11, 16, 17, 20 y 28. Para luego pedir que respondieran las siguientes preguntas:

- En la primera visualización (Visualización 9), ¿qué sentimiento es el más predominante?
- En la segunda visualización (Visualización 11), ¿qué sentimiento es el que más varía?
- En la tercera visualización (Visualización 16), ¿cuál hashtag es el que más se repite?
- En la tercera visualización (Visualización 16), ¿qué sentimiento concentra más hashtags?
- En la cuarta visualización (Visualización 17), ¿cuál hashtag es el que más se repite?
- En la cuarta visualización (Visualización 17), ¿qué sentimiento concentra más hashtags?
- En la quinta visualización (Visualización 20), ¿cuál constituyente hizo una mejor campaña por el Apruebo?
- En la quinta visualización (Visualización 20), ¿cuál constituyente hizo una mejor campaña por el Rechazo?
- En la quinta visualización (Visualización 20), ¿cuántos constituyentes apoyaban al Apruebo?
- En la quinta visualización (Visualización 20), ¿en qué cuadrante ubicarías a los constituyentes que estaban por el Apruebo, pero fueron evaluados negativamente?
- En base a lo visto en la última visualización (Visualización 28), del 1 al 7, ¿qué tan influyente crees que fue la aprobación de la norma que definía a Chile como una nación plurinacional e intercultural en el resultado del plebiscito?

- Del 1 al 7, ¿qué tan clara es la información entregada en la presente visualización? (aplica para cada una)
- Del 1 al 7, ¿qué tan útil es la información que se te presenta en esta visualización? (aplica para cada una)
- Del 1 al 7, ¿qué tan sorprendido/a quedas con la información presentada en la visualización? (aplica para cada una)
- Del 1 al 7, ¿qué tan cercana es la información presentada en la visualización respecto a tus conocimientos previos? (aplica para cada una)

4.1.4. Cuestionario de salida

Una vez que los participantes respondieron el cuestionario de evaluación fue momento para ellos de responder el cuestionario de salida, el cual constó de las mismas preguntas del cuestionario de entrada, con el fin de medir el cambio en la percepción del debate constitucional luego de adquirir la información presente en las visualizaciones enseñadas.

4.2. Resultados de la validación

Luego de que los 31 participantes contestaron todas las preguntas correspondientes, se analizaron sus respuestas para determinar la calidad de las visualizaciones de datos confeccionadas y su impacto.

4.2.1. Cuestionario de entrada

Pregunta	Hombres	Mujeres	Total
Del 1 al 7, indica tu nivel de conocimiento respecto del proceso constitucional del 2022.	5,1	5,4	5,2
Del 1 al 7, ¿qué tan negativa crees que era la percepción de la gente respecto a la Convención Constituyente?	4,7	4,8	4,7
Del 1 al 7, ¿qué tan positiva crees que era la percepción de la gente respecto a la Convención Constituyente?	3,1	3,3	3,2

Tabla 23: Respuestas del grupo de validación en el cuestionario de entrada, segmentado por sexo.

Fuente: Elaboración Propia.

En la Tabla 23, se puede apreciar la percepción que los participantes tenían acerca del proceso constitucional y la convención antes de ver las visualizaciones, donde se puede ver que tanto hombres como mujeres en el grupo de validación se definieron bastante conocedores del tema, promediando un 5,2 de 7. Por otro lado, ambos subgrupos estuvieron de acuerdo en que la percepción de la gente respecto a la Convención Constituyente fue más negativa que positiva, con 1,4 puntos de diferencia. Es importante también notar que la segmentación por sexo del grupo de validación no mostró grandes disimilitudes, dado que los puntajes estuvieron bastante cercanos, aunque se puede concluir que las mujeres se consideraron a sí mismas como más conocedoras del tema que los hombres.

Respecto a las otras dos preguntas, las que no tenían asociadas un puntaje, ambos grupos acordaron que Twitter no es una herramienta ideal para predecir resultados políticos, con sólo un 29,1% de los encuestados manifestándose a favor del uso de esta red social como agente predictor para escenarios políticos. En cuanto a la pregunta restante, al ser de carácter más abierta, se registraron muchas respuestas diferentes, aunque algunas encontraron puntos en común entre los encuestados. Las 3 causas principales del triunfo del Rechazo según el grupo de validación fueron las siguientes:

1. Mal comportamiento de los constituyentes/Mal trabajo (11 respuestas)
2. Contenido del borrador constituyente (7 respuestas)
3. Anti campaña de la derecha chilena (4 respuestas)

4.2.2. Cuestionario de evaluación

Al igual que para el cuestionario de entrada, las preguntas con un puntaje asociado fueron tabuladas de acuerdo a cada visualización.

Pregunta (Visualización 9)	Hombres	Mujeres	Total
Del 1 al 7, ¿qué tan clara es la información entregada en la presente visualización?	6,3	6,1	6,2
Del 1 al 7, ¿qué tan útil es la información que se te presenta en esta visualización?	5,8	6,0	5,9
Del 1 al 7, ¿qué tan sorprendido/a quedas con la información presentada en la visualización?	6,5	6,5	6,5
Del 1 al 7, ¿qué tan cercana es la información presentada en la visualización respecto a tus conocimientos previos?	4,1	3,9	4,0

Tabla 24: Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 9, segmentado por sexo.

Fuente: Elaboración Propia.

La visualización 9 resultó ser muy clara para los participantes del experimento, según puede verse en la Tabla 24, lo cual tiene sentido dado que es un gráfico de barras, una de las visualizaciones más básicas dentro de las que se pueden encontrar, promediando un 6,2 de claridad sobre 7. Sin embargo, el dato importante que se logró recabar es que el grupo de validación quedó bastante sorprendido con la información presentada, lo cual se debió a la diferencia considerable que hubo entre la percepción negativa y positiva, ya que los participantes pensaban que la diferencia se iba a asemejar a los resultados finales del plebiscito, quizás con un mayor grado de holgura, pero nunca tanto como el señalado.

Pregunta (Visualización 11)	Hombres	Mujeres	Total
Del 1 al 7, ¿qué tan clara es la información entregada en la presente visualización?	6,1	6,0	6,0
Del 1 al 7, ¿qué tan útil es la información que se te presenta en esta visualización?	5,6	5,5	5,5
Del 1 al 7, ¿qué tan sorprendido/a quedas con la información presentada en la visualización?	6,2	6,1	6,1
Del 1 al 7, ¿qué tan cercana es la información presentada en la visualización respecto a tus conocimientos previos?	3,8	3,7	3,7

Tabla 25: Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 11, segmentado por sexo.

Fuente: Elaboración Propia.

De la misma manera, la Tabla 25 muestra que la visualización 11 resultó ser clara, útil y sorprendente para los participantes aunque en un grado levemente menor. Es posible apreciar eso sí, que la información mostrada en esta visualización resultó estar más alejada de su conocimiento inicial, cuyo puntaje promedió 3,7.

Pregunta (Visualización 16)	Hombres	Mujeres	Total
Del 1 al 7, ¿qué tan clara es la información entregada en la presente visualización?	5,2	4,9	5,1
Del 1 al 7, ¿qué tan útil es la información que se te presenta en esta visualización?	4,3	3,9	4,1
Del 1 al 7, ¿qué tan sorprendido/a quedas con la información presentada en la visualización?	3,6	4,0	3,8
Del 1 al 7, ¿qué tan cercana es la información presentada en la visualización respecto a tus conocimientos previos?	4,5	4,4	4,4

Tabla 26: Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 16, segmentado por sexo.

Fuente: Elaboración Propia.

La nube de *hashtags* categorizada por sentimiento resultó ser menos clara y útil que las dos visualizaciones anteriores, además que su contenido fue menos sorprendente para los participantes, dado que varios reconocieron que se esperaban dicha distribución de los *hashtags* debido a que los usuarios de derecha “*twittean* más”.

Pregunta (Visualización 17)	Hombres	Mujeres	Total
Del 1 al 7, ¿qué tan clara es la información entregada en la presente visualización?	6,1	6,1	6,1
Del 1 al 7, ¿qué tan útil es la información que se te presenta en esta visualización?	4,4	3,9	4,1
Del 1 al 7, ¿qué tan sorprendido/a quedas con la información presentada en la visualización?	3,6	4,0	3,8
Del 1 al 7, ¿qué tan cercana es la información presentada en la visualización respecto a tus conocimientos previos?	4,5	4,4	4,4

Tabla 27: Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 17, segmentado por sexo.

Fuente: Elaboración Propia.

En cuanto a la utilidad de la información, según la Tabla 27, la percepción se mantuvo igual para el gráfico radial de *hashtags* categorizados por sentimiento, dado que contenían los mismos datos. Sin embargo, la claridad de la información aumentó considerablemente respecto a la visualización desarrollada como nube de palabras, marcando un punto completo más en esta presentación, que resultó ser más intuitiva para los participantes.

Pregunta (Visualización 20)	Hombres	Mujeres	Total
Del 1 al 7, ¿qué tan clara es la información entregada en la presente visualización?	5,5	5,6	5,6
Del 1 al 7, ¿qué tan útil es la información que se te presenta en esta visualización?	6,2	5,9	6,0
Del 1 al 7, ¿qué tan sorprendido/a quedas con la información presentada en la visualización?	4,1	4,3	4,2
Del 1 al 7, ¿qué tan cercana es la información presentada en la visualización respecto a tus conocimientos previos?	5,2	5,0	5,1

Tabla 28: Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 20, segmentado por sexo.

Fuente: Elaboración Propia.

En la Tabla 28 se puede ver que algunas personas tuvieron problemas para entender la visualización 20, pese a que el promedio de claridad fue bastante alto, con un 5,6, mientras que la

utilidad de la información fue bien apreciada por los participantes. El indicador importante en este gráfico, es que los usuarios tenían previsto los posicionamientos de los constituyentes en base a sus conocimientos previos.

Pregunta (Visualización 28)	Hombres	Mujeres	Total
Del 1 al 7, ¿qué tan clara es la información entregada en la presente visualización?	6,0	6,1	6,0
Del 1 al 7, ¿qué tan útil es la información que se te presenta en esta visualización?	6,3	6,3	6,3
Del 1 al 7, ¿qué tan sorprendido/a quedas con la información presentada en la visualización?	6,5	6,7	6,6
Del 1 al 7, ¿qué tan cercana es la información presentada en la visualización respecto a tus conocimientos previos?	3,5	3,9	3,7

Tabla 29: Respuestas del grupo de validación en el cuestionario de evaluación para la Visualización 28, segmentado por sexo.

Fuente: Elaboración Propia.

La última visualización mostrada (Figura 28), que muestra el impacto en el tiempo de la aprobación de la norma que haría de Chile una nación plurinacional e intercultural, fue recibida con bastante sorpresa por parte del grupo de validación, de acuerdo a la Tabla 29, que en una escala del 1 al 7 se sorprendieron en promedio en un 6,6, la cifra más alta registrada en esta evaluación. Cabe destacar también que esta visualización mostró el nivel mayor en utilidad también, con un 6,3. Dicha visualización presentó una pregunta adicional (En base a lo visto en la última visualización (Visualización 28), del 1 al 7, ¿qué tan influyente crees que fue la aprobación de la norma que definía a Chile como una nación plurinacional e intercultural en el resultado del plebiscito?), que no fue incluida en ninguna de las tablas anteriores. Sin embargo, sus puntajes asociados sí fueron registrados, marcando un promedio de 5,7 en los hombres y un 6,0 en las mujeres, con un promedio aproximado de 5,8.

El resto de las preguntas fueron evaluados en base a una tasa de éxito y fracaso, con el fin de medir aspectos más específicos para cada visualización.

Pregunta	Éxito	Fracaso
En la primera visualización (Visualización 9), ¿qué sentimiento es el más predominante?	100 %	0 %
En la segunda visualización (Visualización 11), ¿qué sentimiento es el que más varía?	80,6 %	19,4 %
En la tercera visualización (Visualización 16), ¿cuál hashtag es el que más se repite?	51,6 %	48,4 %
En la tercera visualización (Visualización 16), ¿qué sentimiento concentra más hashtags?	100 %	0 %
En la cuarta visualización (Visualización 17), ¿cuál hashtag es el que más se repite?	87,1 %	12,9 %
En la cuarta visualización (Visualización 17), ¿qué sentimiento concentra más hashtags?	100 %	0 %
En la quinta visualización (Visualización 20), ¿cuál constituyente hizo una mejor campaña por el Apruebo?	74,2 %	25,8 %
En la quinta visualización (Visualización 20), ¿cuál constituyente hizo una mejor campaña por el Rechazo?	83,9 %	16,1 %
En la quinta visualización (Visualización 20), ¿cuántos constituyentes apoyaban al Apruebo?	67,8 %	32,2 %
En la quinta visualización (Visualización 20), ¿en qué cuadrante ubicarías a los constituyentes que estaban por el Apruebo, pero fueron evaluados negativamente?	61,3 %	38,7 %

Tabla 30: Porcentaje de éxito y fracaso obtenido por el grupo de validación
Fuente: Elaboración Propia.

En general, las tareas asignadas a cada participante fueron completadas de manera exitosa, con un porcentaje de acierto que da a entender que las visualizaciones lograron entregar de manera óptima la información y los resultados obtenidos en este estudio. Yendo a un plano más específico, la nube de palabras tuvo problemas para diferenciar los *hashtags* unos de otros en cuanto a tamaño, debido a que la escala no fue la adecuada para hacer la distinción entre éstos, lo cual se solucionó considerablemente con el uso del gráfico radial, cuyo porcentaje de éxito respecto a la misma tarea aumentó considerablemente, pasando de un 51,6 % a un 87,1 %. Por otro lado, el gráfico de dispersión, el cual ubicó a los constituyentes en base a sentimiento y posición en el proceso, no mostró problemas para identificar los puntos representados en los extremos, pero sí le hizo la tarea más complicada a los participantes al momento de determinar agrupaciones más complejas, como las últimas dos presentadas en la Tabla 30.

4.2.3. Cuestionario de salida

Luego de completado el experimento, los participantes respondieron exactamente las mismas preguntas presentadas en el cuestionario de entrada, para luego medir el cambio provocado por la información presentada en las visualizaciones ya evaluadas.

Pregunta	Hombres	Mujeres	Total
Del 1 al 7, indica tu nivel de conocimiento respecto del proceso constitucional del 2022.	5,6	5,8	5,7
Del 1 al 7, ¿qué tan negativa crees que era la percepción de la gente respecto a la Convención Constituyente?	5,9	6,2	6,0
Del 1 al 7, ¿qué tan positiva crees que era la percepción de la gente respecto a la Convención Constituyente?	2,6	2,8	2,7

Tabla 31: Respuestas del grupo de validación en el cuestionario de salida, segmentado por sexo.

Fuente: Elaboración Propia.

Las visualizaciones tuvieron un impacto en la percepción general de los participantes, quienes ahora consideran saber más sobre el proceso, indicando que la percepción negativa del público es más alta de lo que inicialmente creían, mientras que la percepción positiva bajó, lo cual se condice con la información entregada mediante las visualizaciones enseñadas. Al preguntarles nuevamente por la factibilidad de usar Twitter como herramienta para predecir escenarios políticos, la opinión positiva subió, pero no alcanzó a ser mayoría, pasando de un 29,1 % a un 35,5 %. Finalmente, los participantes cambiaron de opinión en cuanto a las 3 causas principales del triunfo del Rechazo, cuyas 3 primeras respuestas se presentan de la siguiente manera:

1. Mal comportamiento de los constituyentes/Mal trabajo (10 respuestas)
2. Contenido del borrador constituyente (6 respuestas)
3. Aprobación de la norma de plurinacionalidad e interculturalidad (5 respuestas)

De esta forma, concluye el proceso de validación de la propuesta de solución, el cual mediante la evaluación de las visualizaciones de datos confeccionadas mostró que la información obtenida fue útil para las personas, que se entregó de manera clara y que logró darle una nueva mirada a la percepción de los participantes del experimento sobre el proceso constitucional de 2022.

CONCLUSIONES

A lo largo de esta investigación, se ha desarrollado un trabajo de periodismo de datos con una metodología basada en KDD que involucró cuatro etapas, cuyo objetivo fue hallar los hechos y tendencias que marcaron el plebiscito constitucional de 2022. Para esto, se decidió utilizar como fuente de datos la red social Twitter, estableciendo un lapso de estudio comprendido entre enero y septiembre de 2022, este último donde, finalmente, se llevó a cabo dicho plebiscito, con el fin de aplicar análisis de sentimiento sobre las opiniones de las personas vertidas en cada tweet, para luego confeccionar diversas visualizaciones de datos que lograran construir una narrativa consistente y coherente acerca de los hallazgos realizados en este trabajo, cuya utilidad, finalmente, fue evaluada por un grupo de personas que tuvieron acceso a dichas visualizaciones y que entregaron su opinión que pudo ser analizada de manera cualitativa y cuantitativa.

Comenzando por la recopilación de datos, se puede concluir que el volumen y lapso considerados para la ejecución de la solución propuesta fueron suficientes para entregar resultados frescos y nuevos acerca del proceso constitucional de 2022, abriendo paso a una mirada desconocida, que escapa a la mayoría de las personas que hablan y opinan sobre el tema. En cuanto a la selección, transformación y limpieza de los datos, es posible afirmar que los atributos seleccionados y los algoritmos diseñados también fueron suficientes para la consecución de los objetivos planteados al inicio del trabajo, dado que el análisis de los resultados entregó una cantidad casi nula de resultados que no correspondieran al contexto de esta memoria, cuyo mayor ejemplo se puede visualizar en las nubes de palabras y gráficos radiales de *hashtags* que, en su mayoría, contienen palabras atinentes al debate constitucional en general.

La interpretación de los datos mediante el uso de análisis de sentimiento, emociones y detección de discurso de odio también puede ser considerada un éxito, tanto en forma como en fondo, dado que las métricas obtenidas mediante el uso de la matriz de confusión, arrojaron que los modelos seleccionados fueron una elección correcta para la tarea, otorgando altos niveles de precisión y exactitud, mientras que la decisión de emplear tres tipos de análisis diferentes basados en el procesamiento del lenguaje natural, aportó positivamente para la construcción de la narrativa que se propuso en un inicio, logrando establecer relaciones causales entre el Rechazo, la percepción negativa de la Convención Constituyente, el enojo y la presencia de discurso de odio, fortaleciendo los resultados y las hipótesis planteadas a lo largo de este trabajo.

Utilizar visualizaciones de datos para finalmente darle forma a la narrativa que se entrega en este documento es un toque que permite no sólo visualizar de forma más clara las tendencias y los hechos que determinaron el resultado del plebiscito, sino que también amplían el alcance de esta memoria, logrando llevar al público información fácil de procesar, permitiendo adquirir nuevo conocimiento acerca de uno de los procesos políticos más importantes ocurridos en nuestro país durante los últimos 20 años.

En resumen, también es válido concluir que el uso de la metodología KDD para llevar a cabo esta investigación fue una decisión correctamente tomada, teniendo en cuenta las herramientas que se deseaban utilizar y los recursos disponibles, cuyo orden lógico de las etapas calzó perfectamente con los objetivos generales y específicos que se propusieron al inicio de esta memoria.

Diferencia entre las encuestas y la percepción obtenida

Al inicio de este documento, se introdujo, en la Figura 1, un gráfico construido en base a la evolución en el tiempo de las encuestas realizadas sobre la posición, Apruebo o Rechazo, respecto al plebiscito constitucional de 2022, durante los meses de enero y septiembre, el mismo lapso empleado para la solución de este estudio. Ahí se puede ver una tendencia positiva para el Apruebo, alcanzando el 60 % de las preferencias, que finalmente caerían estrepitosamente hasta llegar al 38 % final que se obtuvo al momento del plebiscito. Sin embargo, esta realidad dista bastante de la presentada por la narrativa de esta memoria, la cual sostiene que la percepción acerca de la convención fue siempre negativa y jamás estuvo cerca de alcanzar niveles positivos que lograra igualar las tendencias.

La diferencia de resultados entre las encuestas y el trabajo realizado aquí pudo deberse a varios factores, como tamaño de la muestra, estructura de la muestra, sesgo en los datos, errores en la metodología, etc. Sin embargo, la razón principal que puede explicar tan cuantiosa disimilitud radica en el origen de los datos, recordando que el conjunto de datos construido para construir la percepción de la gente en torno a la Convención Constituyente proviene desde una red social, Twitter, la cual al ser un espacio libre de expresión, abre la puerta a una mayor cantidad de comentarios odiosos en general, que abundaron en el *dataset*. La negatividad en redes sociales ha sido tema de estudio en los últimos años, sobre todo Twitter, que según Ferrán Lalueza, profesor de Estudios de Ciencias de la Información y de la Comunicación en la Universitat Oberta de Catalunya, es “una red social eminentemente textual y eso es algo que imprime cierto carácterz que “en general, las plataformas audiovisuales tienen un carácter más amable que las textuales” [Lalueza, 2020], marcando una diferencia con competidores como Instagram, que privilegia el contenido multimedia por sobre el texto, a diferencia de Twitter. Además, Lalueza apunta a que la limitación de extensión en caracteres para cada *tweet* “hace que los mensajes difundidos por medio de ella tiendan a la simplificación y al maniqueísmo que excluye los matices, lo cual a su vez lleva a la polarización y a un enfrentamiento muy radicalizado” [Lalueza, 2020], algo que se reflejó claramente en los múltiples análisis que se llevaron a cabo sobre un *dataset* plagado de *tweets*, quienes tomaron en su mayoría una posición detractora frente a la Convención Constituyente y fueron bastante ácidos en sus críticas.

Sin embargo, analizar el pensamiento más radicalizado de la gente agrega perspectivas desconocidas a estudios sobre escenarios políticos, principalmente encuestas, que en su mayoría no consideran a las redes sociales como fuentes de información, sino que prefieren la vía telefónica, usualmente obteniendo opiniones más moderadas y privilegiando variables de tipo cuantitativas, es decir, no logran capturar el sentimiento de las personas respecto

a temas más específicos como los tratados en este trabajo. Aún así, lo realizado en esta investigación logró detectar una tendencia al alza en la percepción negativa de las personas respecto al proceso constitucional, dado que las curvas de negatividad, enojo y discurso de odio en las Figuras 14, 23 y 26 tienen el mismo comportamiento que el movimiento detectado en el gráfico presentado en la Figura 1, con subidas sostenidas entre los meses de febrero y mayo.

Hechos y tendencias que determinaron el resultado del plebiscito constitucional de 2022

El objetivo general y título de esta memoria trata sobre encontrar los hechos y tendencias que finalmente decantaron en la victoria del Rechazo el día 4 de septiembre de 2022, derrotando a un Apruebo que según las encuestas había comenzado como la opción fuerte, pero que conforme avanzaba el 2022 rápidamente cedió terreno hasta el resultado que hoy se conoce en la actualidad. El trabajo realizado en esta memoria logró sintetizar en tres grandes ejes las causales de la victoria del Rechazo, de acuerdo al sentimiento general de los votantes en Twitter.

Presencia en Twitter de la derecha chilena y el Rechazo A lo largo de este estudio, se determinó que la percepción negativa de la Convención Constituyente fue bastante mayor a la esperada en un inicio y superó con creces a la percepción positiva. Uno de los motivos principales fue la mejor presencia de usuarios de derecha y pro Rechazo en Twitter, quienes empujaron con mayor fuerza sus ideas, logrando influenciar a una amplia cantidad de personas conforme avanzaba el 2022. Durante la investigación, se descubrió que la mayoría de los *hashtags* utilizados durante el periodo estudiado eran relativos al Rechazo, y que en su mayoría se encontraban ligados a un sentimiento negativo, como se puede ver en la Figura 17, lo cual se traduce en un mayor poder de llegada hacia los demás *twitteros*, quienes fueron mejor influenciados por este punto de vista que por el Apruebo.

También, al hacer un análisis en cuanto a los usuarios que participaron en el debate constitucional se identificó que las personas que más *twittearon* durante el periodo eran personas que correspondían a un perfil radicalizado de derecha, con ideas “anticomunistas”, consigna que se repitió constantemente en *tweets* de connotación negativa, cuyo ejemplo más claro puede verse en el *hashtag* #RechazoElMamarrachoComunista, haciendo referencia al borrador constitucional.

Por último, el análisis de los personajes más mencionados también apoya la idea de que la derecha y el Rechazo tuvieron un mayor grado de influencia que el Apruebo, como se puede ver en las Figuras 20 y 21, con una derecha actuando en bloque en ambos casos, con sus militantes ubicándose casi perfectamente en las esquinas inferiores izquierdas, a diferencia de sus pares de izquierda, que se dispersaron más en sus respectivas áreas, dando a entender que no lograron el mismo poder de coordinación y por ende, perdieron el arrastre necesario para influenciar a los votantes a sufragar por el Apruebo.

Rechazo al gobierno de Gabriel Boric El análisis de los personajes más mencionados durante el periodo estudiado arrojó que el actual presidente de la república, Gabriel Boric, fue el no constituyente más citado en este lapso, lo cual pudo haber sido bastante positivo para el Apruebo, dado que Boric pertenece a uno de los bandos que más campaña hizo por este sector, y además fue uno de los principales impulsores del acuerdo constitucional que originó este proceso, en el año 2019, para superar la crisis política que se vivía en Chile en ese entonces. Sin embargo, la presencia del presidente en redes sociales distó mucho de lo que realmente se esperaba, atrayendo en su mayoría opiniones negativas respecto a la Convención Constituyente. Cabe preguntarse entonces, ¿por qué ocurrió esto?.

La respuesta se encuentra en que una buena porción de los usuarios de Twitter que comentaron en el debate vieron a Boric como un personaje que intentó influenciar a la población a votar por el Apruebo desde su posición de presidente, y que además, lo acusaron de basar su proyecto de gobierno en la victoria de esta misma opción. Esto, junto con las polémicas que ha sufrido su gobierno y una creciente tasa de desaprobación en encuestas (véase Figura 29), provocaron que los *twitteros* comenzaran a asociar al gobierno de Gabriel Boric con el Apruebo, traduciendo esta desaprobación a un crecimiento de la percepción negativa de la convención, es decir, sumando adeptos por el Rechazo.

Analizando los *hashtags* presentes en el debate, donde aparecieron frases como #BoricDictador y #GobiernoDeInutiles, evidencian lo expuesto en el párrafo anterior, con usuarios atacando al gobierno de Boric al mismo tiempo que opinaban negativamente sobre la Convención Constituyente, ya que es importante señalar, que estos *hashtags* aparecieron en un conjunto de datos que contiene exclusivamente *tweets* sobre dicho organismo redactor, y no sobre el gobierno ni el presidente, cuya mera presencia refuerza nuevamente la idea de que los usuarios de Twitter vieron a la convención y al gobierno como símiles.

Plurinacionalidad e interculturalidad El último eje presentado por este estudio guarda relación con la aprobación de la norma que consagraba a Chile como un estado regional, plurinacional e intercultural, hecho que, como se puede ver en la Figura 28, provocó un aumento en el discurso de odio hacia la Convención Constituyente y, en consecuencia, empeoró la percepción de las personas en torno a este organismo, que fue acusado de tener intenciones “refundacionales” para nuestro país.

La aprobación de dicha norma despertó comentarios conservadores de los *twitteros*, quienes comenzaron a comentar con *hashtags* como #RechazoXAmorAChile y #RechazoLaDestruccionDeChile, dando a entender que votar por el Rechazo era una forma de defender a Chile de los cambios propuestos por la Convención Constituyente. En Chile, las tendencias conservadoras se encuentran estrictamente ligadas a la derecha política, como apunta Hugo Herrera, filósofo chileno, en su libro “La derecha en la crisis del bicentenario”, quien dice que en la derecha chilena “se distingue una tradición cristiana y liberal -moralmente conservadora, pero vinculada en lo económico a nociones como el librecambismo, el capitalismo y últimamente la subsidiariedad negativa” [Herrera, 2014].

De esta manera, los ataques a la misma norma y al “nuevo Chile” que supuestamente la

Convención Constitucional quería fundar se empezaron a hacer notar con más fuerza y odio, en *hashtags* como #RechazoElPlurimamarracho, y llamando a quienes se mostraban a favor de esta norma, argumentando que sería un cambio histórico para nuestro país como “pluriweones”, lo cual constituye un insulto directo hacia dichas personas.

Validación de las visualizaciones

Como se vio en el capítulo 4, los participantes del experimento de validación de la solución propuesta por estudio evaluaron de buena manera las visualizaciones que resultaron de éste, calificándolas como claras en la forma de entregar la información y útil el conocimiento adquirido, además que también se mostraron sorprendidos por los resultados, dado que varios distaban de lo que ellos originalmente creían.

Sin embargo, visualizaciones como las nubes de palabras o el mapa que ubicaba a los personajes públicos, a pesar de tener un buen desempeño, fueron difíciles de comprender para una parte considerable del grupo de validación, quienes tuvieron serias dificultades para identificar detalles más específicos dentro de ellas, por ende, no tuvieron el rendimiento esperado para esta memoria, pese a lograr su objetivo.

Trabajo futuro

El final de una investigación muchas veces puede ser el comienzo de otra, por lo cual esta memoria busca inspirar a nuevos memoristas e investigadores a continuar en el trabajo basado en el análisis del lenguaje natural, que con el exponencial crecimiento de las tecnologías cada vez tendrá un mejor rendimiento y precisión.

En este trabajo, se deja propuesto ampliar el conjunto de datos a trabajar, yendo de Twitter a otras redes sociales como Facebook o Instagram, que pueden introducir a una mirada más amigable de tópicos tan duros como los procesos políticos, según la mirada de Lalueza. De la misma manera, también podría ampliarse el rango de datos estudiados, y considerar desde la conformación del organismo redactor en el año 2021, o incluso desde su génesis en el acuerdo constitucional de 2019.

Finalmente, el uso de nuevas técnicas o modelos especializados también se deja propuesto en esta memoria, ya que podría entregar nuevos *insights* sobre el sentimiento generalizado de las personas en redes sociales, estableciendo correlaciones más interesantes que las halladas en este estudio dando origen, de esta forma, a visualizaciones de datos que capturen con mayor fuerza y entusiasmo el interés de la gente por el uso de la informática como una herramienta para ayudar a describir procesos políticos.

ANEXOS

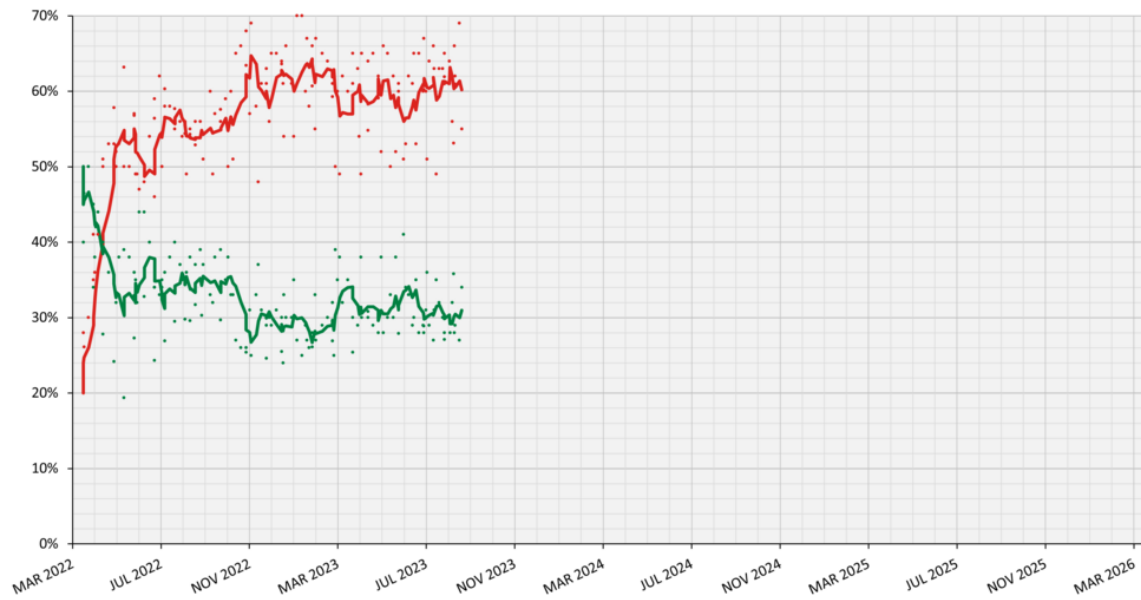


Figura 29: Evolución de la aprobación de Gabriel Boric durante el año 2022.

Fuente: Elaboración propia.

REFERENCIAS BIBLIOGRÁFICAS

- [Asale y Rae, 2020] Asale, R. y Rae (2020). Constitución: Diccionario de la lengua española. <https://dle.rae.es/constituci%C3%B3n>.
- [BCN, 2022] BCN (2022). Con histórica participación electoral propuesta de nueva constitución fue rechazada.
- [Crucianelli, 2013] Crucianelli, S. (2013). ¿qué es el periodismo de datos? *Issuu*.
- [Espín-Riofrío et al., 2021] Espín-Riofrío, C., Zumba Gamboa, J., y Cruz Chóez, A. (2021). Métodos de extracción de comentarios de la red social twitter para uso en procesamiento de lenguaje natural. *Métodos de extracción de comentarios de la red social Twitter para uso en Procesamiento de Lenguaje Natural*.
- [Herrera, 2014] Herrera, H. (2014). *La derecha en la crisis Del Bicentenario*. Ediciones Universidad Diego Portales.
- [Lalueva, 2020] Lalueva, F. (2020). ¿por qué twitter es la red del odio? <https://www.uoc.edu/portal/es/news/actualitat/2020/165-twitter-red-odio.html>.
- [Sahayak et al., 2015] Sahayak, V., Shete, V., y Pathan, A. (2015). Sentiment analysis on twitter data. *Academia.edu*.
- [Sarlan y Basri, 2014] Sarlan, A. y Basri, S. (2014). *Twitter Sentiment Analysis*.
- [Senado, 2019] Senado (2019). Logran histórico acuerdo para nueva constitución: Participación ciudadana será clave - senado - república de chile.
- [Stanford, 2012] Stanford (2012). Stanford classifier. <https://nlp.stanford.edu/wiki/Software/Classifier>.