

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
CIUDAD - CHILE



“ANÁLISIS DE LA PERCEPCIÓN HACIA PUBLICACIONES
SOBRE SALUD MENTAL EN UNA RED SOCIAL EN CHILE
MEDIANTE TÉCNICAS DE MACHINE LEARNING Y
VISUALIZACIÓN DE DATOS”

DEBORA NICOLE ALAYO AZABACHE

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL EN INFORMÁTICA

Profesor Guía: José Luis Martí Lara
Profesor Correferente: Mauricio Olivares

Agosto - 2025



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): Memoria o trabajo de título; Tesis de Postgrado;

Título del trabajo: Análisis de la percepción hacia publicaciones sobre salud mental en una red social en Chile mediante técnicas de machine learning y visualización de datos

Nombre del candidato(a): Debora Nicole Alayo Azabache

Carrera / Grado: Ingeniería Civil Informática

Campus: Santiago San Joaquín ; **Departamento:** Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, José Luis Martí Lara , en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

El trabajo **NO contiene información que amerite confidencialidad** y puede ser publicado de inmediato en repositorio con acceso abierto.

El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

6 meses; 12 meses; 2 años; 3 años; 5 años; 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha:
23 - 08 - 2025

; Firma:

Estudiante o Candidato(a):

Fecha: 22-08-2025

; Firma:

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

DEDICATORIA

A mi mamá.

AGRADECIMIENTOS

Quiero agradecer a mi amada familia, a mi hermosa madre que es la mujer más increíble que conozco: Jenny, sin ella jamás habría llegado hasta aquí. A mis hermanos menores: Junior, Jenniffer y Antonio que son lo más importante para mí. A mis bellas perritas: Becky y Kuki, que son las bendiciones más locas y brillantes que me ha dado la vida. A mi abuela Esther y a Jacobito que me dieron su apoyo cuando lo necesité. A la segunda familia que me regaló esta universidad: Alonso, Alan, Varo, Pancho, Diego, Joaco, Jose, Cata y Nandy, sin ustedes este camino no habría sido lo lindo que fue. Y lo mejor que me pasó en esta gran etapa, el amor de mi vida, Simón. Te amo demasiado, muchas gracias por todo lo que eres para mí. A todas estas personas, muchas gracias por ser parte de mi vida. Y finalmente, gracias a mi versión más pequeña, a aquella niña soñadora, fuerte y alegre. Por ti soy todo lo que soy.

RESUMEN

Resumen— Esta memoria aborda el análisis de la percepción de los usuarios en Chile frente a publicaciones sobre salud mental en la red social Instagram. El objetivo principal fue identificar las reacciones emocionales predominantes ante este tipo de contenido, considerando su impacto social y el contexto actual del país. Para ello, se recopiló un conjunto de publicaciones y comentarios mediante técnicas de extracción de datos (*web scraping*), y se aplicaron modelos de análisis de sentimientos y emociones utilizando herramientas de procesamiento de lenguaje natural. La metodología seguida fue CRISP-DM, lo que permitió estructurar el trabajo en fases iterativas de análisis y validación. Los resultados mostraron que las reacciones positivas y de alegría son las más frecuentes, mientras que las emociones negativas, como tristeza o ira, se vinculan a temáticas sociales sensibles, como el género o la discriminación. Las visualizaciones generadas permiten comprender mejor cómo se expresa y percibe la salud mental en el entorno digital, entregando información valiosa para el diseño de intervenciones, campañas educativas o investigaciones futuras.

Palabras Clave— salud mental; redes sociales; análisis de sentimientos; visualización de datos.

ÍNDICE DE CONTENIDOS

RESUMEN	IV
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	VIII
INTRODUCCIÓN	1
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	2
1.1 Contexto	2
1.1.1 La importancia de la perspectiva en la Salud Mental	2
1.1.2 El impacto del estigma en la salud mental	3
1.2 Objetivos	4
CAPÍTULO 2: MARCO CONCEPTUAL	5
2.1 Metodologías	5
2.2 <i>Web Scraping</i>	8
2.3 Visualizaciones de datos	9
2.4 Periodismo de datos	10
2.5 Análisis de Sentimientos	11
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	13
3.1 Selección de la metodología de trabajo	13
3.2 Compresión del negocio	13
3.2.1 Determinación de los objetivos del negocio	13
3.2.2 Evaluación de la situación	14
3.3 Comprensión de los datos	15
3.3.1 Contextualización: Estadísticas sobre salud mental en Chile	15
3.3.2 Selección de la plataforma para el análisis	19
3.3.3 Identificación de <i>hashtags</i> relevantes	20
3.3.4 Extracción de datos: desafíos y soluciones	20
3.3.5 Descripción de los datos	21
3.4 Preparación de los datos	29
3.5 Modelado: primera iteración	30
3.5.1 Resultados	32
3.6 Evaluación: primera iteración	35
3.6.1 Métricas de rendimiento	36
3.6.2 Resultados: Análisis de Sentimientos	42
3.6.3 Resultados: Análisis de Emociones	43
3.7 Modelado: segunda iteración	44
3.7.1 Corrección heurística de emociones a partir de emojis	44
3.8 Evaluación: segunda iteración	47
3.9 Despliegue	49

3.9.1	¿Cómo varía la percepción de los usuarios en el tiempo?	50
3.9.2	¿Qué temas tienden a generar reacciones predominantemente positivas?	53
3.9.3	¿Cómo se expresan las reacciones positivas?	55
3.9.4	¿Qué temas tienden a generar reacciones predominantemente negativas?	58
3.9.5	¿Cómo se expresan las reacciones negativas?	59
3.9.6	¿Qué temáticas están predominantemente vinculadas a la emoción <i>joy</i> ?	60
3.9.7	¿Qué temáticas están predominantemente vinculadas a la emoción <i>anger</i> ?	61
3.9.8	¿Qué temáticas están predominantemente vinculadas a la emoción <i>sadness</i> ?	62
3.9.9	¿Cómo se expresan en los comentarios cada emoción?	63
3.9.10	¿Qué información de valor puede extraerse a partir del uso de emojis en los comentarios?	64
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN		66
4.1	Metodología de la encuesta	66
4.2	Diseño de la encuesta	67
4.2.1	Sección 1: Publicaciones (Sentimientos)	67
4.2.2	Sección 2: Publicaciones (Emociones)	68
4.2.3	Sección 3: Comentarios	69
4.2.4	Sección 4: Preguntas generales y de perfil	71
4.3	Análisis de las respuestas	72
4.3.1	Con respecto a las visualizaciones de palabras más frecuentes.	72
4.3.2	Sobre las visualizaciones temporales.	72
4.3.3	Sobre las visualizaciones de nubes de palabras.	73
4.3.4	Sobre la visualización de coocurrencia de emojis.	73
4.3.5	Sobre la relación entre temáticas sociales y salud mental.	73
4.3.6	Sobre quienes respondieron.	74
4.4	Conclusión de la Validación	74
5 CONCLUSIONES		75
REFERENCIAS BIBLIOGRÁFICAS		80

ÍNDICE DE FIGURAS

1	Representación de las etapas de KDD.	5
2	Representación de las etapas de SEMMA.	6
3	Representación de las etapas de CRISP-DM.	7
4	Proceso de Web Scraping.	8
5	Visualización sobre la conformación del Consejo Constitucional electo en mayo de 2023.	11
6	Mapa de calor de las publicaciones por mes y año para el <i>hashtag</i> SaludMentalChile.	21
7	Mapa de calor de las publicaciones por mes y año para el <i>hashtag</i> PsicologíaChile.	22
8	Mapa de calor de las publicaciones por mes y año para los <i>hashtags</i> SaludMentalChile y PsicologíaChile.	23
9	Matriz de confusión correspondiente al modelo de análisis de sentimientos <i>pysentimiento</i> evaluado sobre una muestra balanceada por clase real.	37
10	Matriz de confusión correspondiente al modelo de análisis de sentimientos de la API de <i>Google Natural Language</i> evaluado sobre una muestra balanceada por clase real.	38
11	Frecuencia de las emociones clasificadas por <i>pysentimiento</i>	43
12	Comparación de la distribución de emociones detectadas por el modelo original y el ajustado con heurísticas.	47
13	Evolución mensual de los sentimientos en los comentarios, clasificados automáticamente como positivos, neutros o negativos mediante la biblioteca <i>pysentimiento</i>	50
14	Mapa de calor que representa la cantidad de comentarios con sentimiento positivo, distribuidos por día de la semana y hora del día.	52
15	Mapa de calor que representa la cantidad de comentarios con sentimiento negativo, distribuidos por día de la semana y hora del día.	52
16	Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios positivos.	54
17	Nube de palabras generada a partir de los comentarios clasificados con sentimiento positivo.	55
18	Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios negativos.	58
19	Nube de palabras generada a partir de los comentarios clasificados con sentimiento negativos.	59
20	Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios con emoción <i>joy</i>	60
21	Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios con emoción <i>anger</i>	61
22	Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios con emoción <i>sadness</i>	62
23	Palabras más frecuentes asociadas a cada emoción.	63

24	Grafo de coocurrencia de emojis utilizados en los comentarios.	65
----	--	----

ÍNDICE DE TABLAS

1	Descripción de los atributos de un <i>post</i> obtenido desde la API utilizada. . . .	24
2	Descripción de los atributos de un comentario en la estructura de datos. . . .	25
3	Descripción de la estructura de comentarios jerárquicos asociados a publicaciones.	26
4	Descripción de los atributos de un comentario y sus posibles respuestas (estructura jerárquica).	27
5	Hashtags más frecuentes en el dataset obtenido.	27
6	Hashtags más frecuentes en publicaciones con el <i>hashtag</i> SaludMentalChile.	28
7	Hashtags más frecuentes en el dataset del <i>hashtag</i> PsicologíaChile.	28
8	Cantidad de comentarios asociados a los <i>hashtags</i> analizados.	29
9	Cabecera del nuevo archivo CSV para realizar el análisis de sentimientos. . . .	29
10	Etiquetas de sentimiento y su significado según el modelo <i>pysentimiento</i>	32
11	Cabecera del archivo CSV que guarda los resultados del análisis de sentimientos hecho con <i>pysentimiento</i>	33
12	Cabecera del archivo CSV que guarda los resultados del análisis de sentimientos hecho con la API de <i>Google Natural Language</i>	34
13	Cabecera del archivo CSV que guarda los resultados del análisis de emociones hecho con <i>pysentimiento</i>	35
14	Comparación entre las clasificaciones de sentimientos de Google y <i>pysentimiento</i> (umbral 0.5)	36
15	Comparación entre las clasificaciones de sentimientos de Google y <i>pysentimiento</i> (umbral 0.25)	36
16	Matriz de confusión - <i>pysentimiento</i>	37
17	Matriz de confusión - <i>Google Natural Language</i>	38
18	Métricas de evaluación - <i>pysentimiento</i>	39
19	Métricas de evaluación - <i>Google Natural Language</i>	39
20	Evolución mensual de los sentimientos en los comentarios, clasificados mediante <i>pysentimiento</i>	41
21	Distribución porcentual mensual de los sentimientos en los comentarios, clasificados mediante <i>pysentimiento</i>	42
22	Distribución de emociones detectadas	43
23	Frecuencia y emoción asignada a cada emoji para los 15 más frecuentes. . . .	46
24	Distribución de emociones detectadas tras segunda iteración	47
25	Ejemplos de comentarios cuya clasificación de emoción fue modificada por la heurística.	48

INTRODUCCIÓN

La salud mental se ha consolidado como una dimensión esencial del bienestar humano, siendo reconocida por organismos internacionales como un componente fundamental para el desarrollo individual y social. En el contexto chileno, diversos estudios han evidenciado un aumento sostenido de los problemas de salud mental en la población, situación que se ha visto intensificada por factores sociales, económicos y sanitarios. A pesar de ello, persisten altos niveles de estigmatización, desinformación y brechas en el acceso a atención especializada. En este escenario, comprender cómo la sociedad percibe y se relaciona con la salud mental es un desafío urgente y necesario.

Paralelamente, el auge de las redes sociales ha transformado profundamente los modos de interacción y comunicación. Instagram, en particular, se ha posicionado como una de las plataformas más utilizadas en Chile, permitiendo la circulación de contenidos sobre salud mental y generando espacios de expresión emocional abiertos y espontáneos. Estas interacciones digitales ofrecen una oportunidad inédita para analizar percepciones colectivas sobre temáticas sensibles, mediante el estudio del lenguaje, los comentarios y otros indicadores expresivos.

Este trabajo se enmarca en el cruce entre el análisis de redes sociales, el procesamiento de lenguaje natural y la visualización de datos, con el objetivo de analizar la percepción de los usuarios chilenos ante publicaciones sobre salud mental en Instagram. Para ello, se propone una solución que combina técnicas de extracción de datos (web scraping), análisis de sentimientos y emociones mediante modelos de aprendizaje automático, y representaciones gráficas que permiten comunicar los hallazgos de manera comprensible.

La metodología adoptada para este estudio corresponde a CRISP-DM, la cual estructura el proceso en fases iterativas que van desde la comprensión del problema y los datos, hasta el modelado, la evaluación y el despliegue de resultados. Esta elección metodológica permite abordar el problema de forma sistemática y adaptable, facilitando ajustes en función de los hallazgos obtenidos.

La presente memoria se estructura de la siguiente forma: en el Capítulo 1 se contextualiza el problema y se detallan los objetivos; el Capítulo 2 presenta el marco conceptual que sustenta el estudio; el Capítulo 3 describe la propuesta de solución, incluyendo la metodología aplicada y el desarrollo técnico; el Capítulo 4 expone el proceso de validación de los resultados con profesionales del área de la psicología. Finalmente, el Capítulo 5 entrega las conclusiones, alcances, limitaciones y posibles líneas de trabajo futuro.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

1.1. Contexto

La salud mental es un tema de vital importancia en una sociedad, pues “es un estado de bienestar mental que permite a las personas hacer frente a los momentos de estrés de la vida, desarrollar todas sus habilidades, poder aprender y trabajar adecuadamente y contribuir a la mejora de su comunidad” [World Health Organization, 2022b]. Sin este estado de bienestar, los pilares de una sociedad son frágiles, pues sus individuos no están en condiciones de dar lo mejor de sí. Este problema se puede apreciar claramente en el aumento de licencias médicas por temas de salud mental, que representa la mayor alza en licencias, específicamente “el aumento de las enfermedades de salud mental en 2020, las que representan un 28,7 % del total (1.730.263) de licencias médicas en Chile, mientras que el año anterior, en 2019, solo representaban el 23,6 % del total” [Superintendencia de Seguridad Social, 2021], esto afecta directamente la fuerza laboral del país.

Y sin duda afecta a la sociedad transversalmente, no solo en el ámbito laboral. Y si bien son las autoridades del Estado sobre quienes recae la mayor responsabilidad de abordar el problema a nivel país, es esencial la propia perspectiva de cada persona, pues esta define la visión con que aborda el problema. Y esto es primordial cuando los recursos externos son escasos, como los que puede entregar el Estado, donde “la cobertura de atención en salud mental en Chile alcanza aproximadamente a no más de un 20 % de la población, mientras que en países de medianos ingresos la cobertura llega a un 50 %” [Centro de Investigación Periodística, 2021].

1.1.1. La importancia de la perspectiva en la Salud Mental

La importancia de la percepción sobre la salud mental radica en el hecho de que nuestra forma de ver y comprender este aspecto influye directamente en cómo nos relacionamos con él, así como en el bienestar general de las personas.

Es fundamental tener una percepción adecuada de la salud mental para poder identificar y reconocer los problemas o trastornos mentales. Si se tiene una percepción adecuada, se puede reconocer cuando algo no está bien y tomar medidas para abordarlo.

Además, una buena percepción de la salud mental permite combatir el estigma asociado a los trastornos mentales. Al entender que los problemas mentales son tan válidos y reales como cualquier otra enfermedad física, se puede promover una cultura de inclusión y apoyo.

Asimismo, la percepción sobre la salud mental también afecta cómo tratamos a quienes padecen trastornos mentales. Si comprendemos que todos tenemos altibajos emocionales en

algún momento de nuestras vidas y que necesitamos apoyarnos mutuamente, podemos fomentar ambientes más empáticos y solidarios.

Tener una buena percepción sobre la salud mental también considera valorar la importancia de cuidar nuestra propia salud mental. Esto implica reconocer que todos tenemos necesidades emocionales y psicológicas, y que debemos tomar medidas para mantener un equilibrio en nuestras vidas. Al igual que cuidamos nuestro cuerpo a través de una alimentación adecuada y ejercicio físico, es fundamental cuidar nuestra mente.

Dada la importancia de la percepción sobre la salud mental, es fundamental conocer cuáles son las percepciones que tiene la sociedad en Chile, pues este conocimiento puede contribuir a generar políticas públicas que eduquen sobre los puntos más estigmáticos y por ende apoyar en el empoderamiento de las personas sobre su propia salud mental.

1.1.2. El impacto del estigma en la salud mental

Lamentablemente, persiste un amplio desconocimiento y prejuicio hacia las personas que sufren problemas mentales, lo que a menudo desencadena discriminación. En ocasiones, debido a estigmas arraigados o a la falta de información, las personas pueden pasar por alto o minimizar los síntomas, evitando buscar ayuda profesional. Si se percibe la salud mental como algo negativo o peligroso, es probable que se evite interactuar con quienes padecen problemas de este tipo o brindarles menos apoyo emocional.

El estigma impacta en todos los aspectos de los problemas de salud mental, siendo más preocupantes las dificultades para reconocer el problema, buscar ayuda, comenzar y completar un tratamiento. También afecta a los seres queridos de quienes padecen problemas mentales, generando sentimientos de desesperanza o temor al no saber cómo ayudar o anticipando cómo serán tratados por otras personas al conocer su situación.

Los efectos negativos del estigma no tienen límites de edad. Puede influir en los niños, quienes podrían malinterpretar la falta de control emocional como simple rebeldía en lugar de considerar la posibilidad de un problema mental subyacente. En adolescentes, las conductas autolesivas o estados de tristeza pueden ser percibidos erróneamente como meros intentos de llamar la atención. En adultos, el cansancio crónico a veces se asume como algo natural, mientras que en la tercera edad, el sufrimiento puede pasar desapercibido y sus señales ser ignoradas, incluso llegando en los casos más graves al suicidio. En Chile, 360 adultos mayores se quitan la vida cada año [Ministerio de Salud, 2019].

Es común también creer que quienes enfrentan problemas de salud mental son personas débiles, que solo necesitan fuerza de voluntad para recuperarse o que cualquier individuo con problemas mentales está "loco". Esto lleva a que muchas personas intenten resolver sus problemas por sí mismas, lo que puede empeorar los síntomas o generar nuevos, prolongando el sufrimiento que, con la ayuda profesional adecuada, podría ser más llevadero y tendría

menos impacto en la calidad de vida.

Incluso en los centros de atención primaria de salud en Chile, el estigma asociado a los trastornos de salud mental continúa representando un desafío significativo, siendo un problema que actualmente se encuentra en proceso de reconocimiento y abordaje progresivo por parte de los equipos de salud [Escuela de Medicina de la Pontificia Universidad Católica de Chile, 2023].

Comprender el estigma y la desinformación que afecta a las personas es fundamental para disminuir su impacto negativo, y evitar que continúe obstaculizando el acceso de las personas a la ayuda profesional que realmente necesitan.

Dado esta necesidad, se proponen los siguientes objetivos para conocer y entender las perspectivas.

1.2. Objetivos

Objetivo general

- Desarrollar un análisis de las reacciones que tienen los usuarios de una red social ante publicaciones de salud mental en Chile mediante técnicas de *machine learning* y periodismo de datos, para conocer la percepción de los usuarios y contribuir a la comprensión del tema.

Objetivos específicos

- Construir un conjunto de datos representativo de publicaciones recientes en redes sociales relacionadas con la salud mental en Chile, incluyendo tanto las publicaciones como los comentarios asociados.
- Aplicar técnicas de *machine learning*, específicamente análisis de sentimientos, sobre los datos recopilados, con el objetivo de identificar las percepciones de los usuarios en torno a la salud mental.
- Desarrollar un proceso de periodismo de datos a partir de la información obtenida, generando visualizaciones que aporten valor a la interpretación y comprensión del contexto investigado.

CAPÍTULO 2

MARCO CONCEPTUAL

2.1. Metodologías

Para realizar el trabajo propuesto es necesario contar con una metodología de trabajo ampliamente utilizada y validada por lo cual se exploran las cuatro más usadas para trabajos de este estilo: CRISP-DM, KDD, SEMMA y Catalyst, que proporcionan estructuras o pasos para guiar el proceso de análisis de datos, desde la identificación del problema hasta la implementación de soluciones.

KDD (Knowledge Discovery in Databases): Trata de un enfoque más amplio que abarca todo el proceso de descubrimiento del conocimiento a partir de grandes volúmenes de datos. KDD consta de cinco etapas principales: selección, preprocesamiento, transformación, minería y evaluación. Ofrece una orientación general sobre las actividades a llevar a cabo en cada etapa sin profundizar en mayor detalle sobre las tareas a ejecutar [Moine *et al.*, 2011].

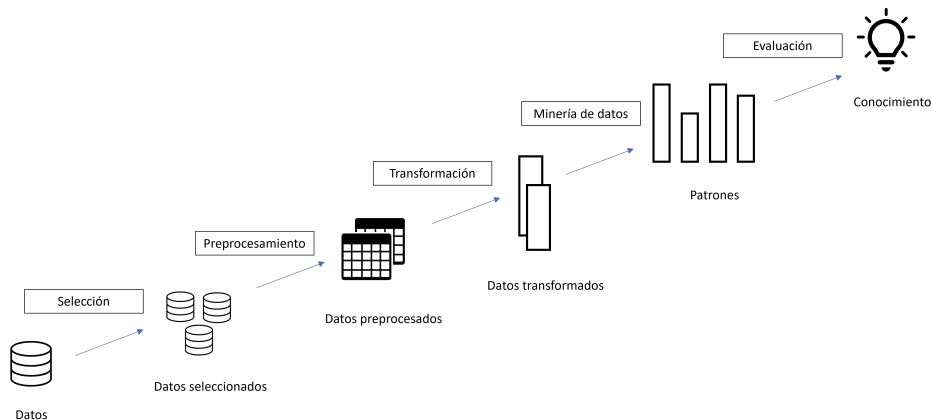


Figura 1: Representación de las etapas de KDD.
Fuente: Elaboración propia.

SEMMA (Sample, Explore, Modify, Model and Assess): Se centra en el análisis predictivo y se utiliza con frecuencia en proyectos relacionados con aprendizaje automático y estadísticas avanzadas. SEMMA sigue un flujo lógico donde primero se toma una muestra representativa del conjunto completo de datos disponibles. Luego, se exploran esos datos para obtener una mejor comprensión antes de realizar modificaciones o transformaciones necesarias. A continuación, se construyen modelos predictivos y, finalmente, se evalúan para medir su eficacia. Se enfoca en aspectos técnicos, dejando de lado las actividades de análisis y comprensión del problema que se está tratando [Moine *et al.*, 2011].

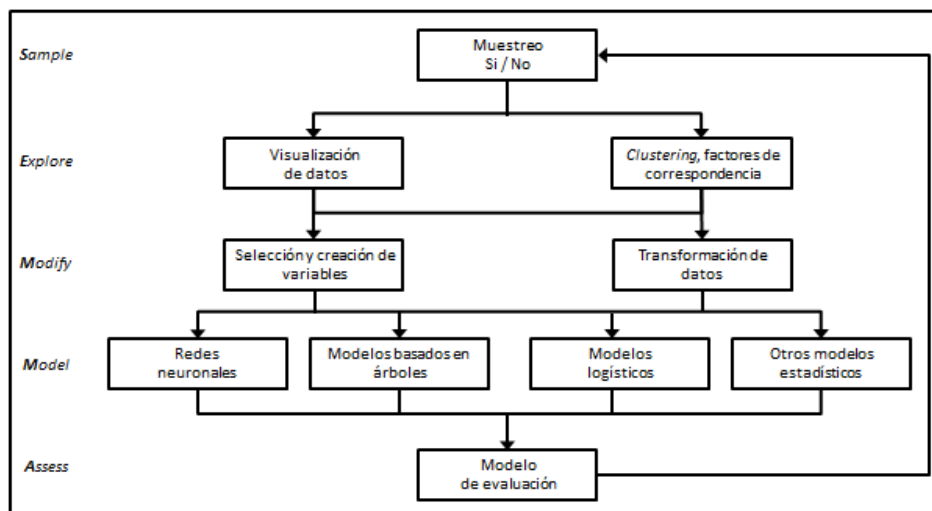


Figura 2: Representación de las etapas de SEMMA.

Fuente: Elaboración propia.

CRISP-DM (Cross-Industry Standard Process for Data Mining): Ampliamente utilizada en proyectos de minería de datos. Fue desarrollada por IBM. Consta de seis fases que se ejecutan iterativamente: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Cada fase tiene sus propias actividades y tareas específicas que ayudan a garantizar un análisis efectivo y una implementación exitosa. CRISP-DM define una serie de tareas y actividades para cada etapa del proyecto, pero no detalla el método o la manera específica de realizarlas [Moine *et al.*, 2011].

A continuación se profundiza en las actividades que puede incluir cada etapa¹.

1. **Comprensión del negocio:** Es la fase inicial donde se busca entender profundamente las necesidades y objetivos del cliente desde una perspectiva de negocio. Implica definir claramente las metas comerciales, evaluar recursos, riesgos y beneficios, establecer objetivos técnicos específicos para la minería de datos, y planificar las herramientas y tecnologías que se utilizarán. Esta etapa es crítica para sentar bases sólidas en el proyecto.
2. **Comprensión de Datos:** En esta fase se identifican, recopilan y analizan los datos necesarios. Incluye obtener los datos iniciales, describir sus propiedades generales, explorarlos detalladamente mediante visualización o consultas.
3. **Preparación de Datos:** Consiste en seleccionar datos relevantes, limpiarlos para corregir errores, crear atributos derivados que aporten valor, integrar datos provenientes de distintas fuentes y formatearlos adecuadamente para su posterior modelado.

¹What is CRISP DM?, Data Science PM. Disponible en: <https://www.datascience-pm.com/crisp-dm-2/>

4. Modelado: Aunque breve, es la fase clave en la ciencia de datos. Aquí se puede seguir el enfoque de desarrollar nuevos modelos o se seleccionan modelos preentrenados adecuados al problema planteado. Posteriormente, se aplican los modelos existentes a los datos.
5. Evaluación: En esta etapa se evalúan los modelos desde una perspectiva empresarial más amplia. Se determina si los modelos satisfacen los criterios comerciales establecidos, se revisa la integridad del proceso seguido y se decide si avanzar con la implementación, iterar más o comenzar nuevos proyectos.
6. Despliegue: Es la fase final y puede variar desde entregar un simple reporte hasta implementar sistemas complejos de minería de datos. Incluye elaborar un plan de implementación, definir estrategias de seguimiento y mantenimiento, producir informes finales para documentar resultados y realizar revisiones del proyecto para futuras mejoras. Se enfatiza que un modelo desplegado debe ser continuamente monitoreado y actualizado según sea necesario.

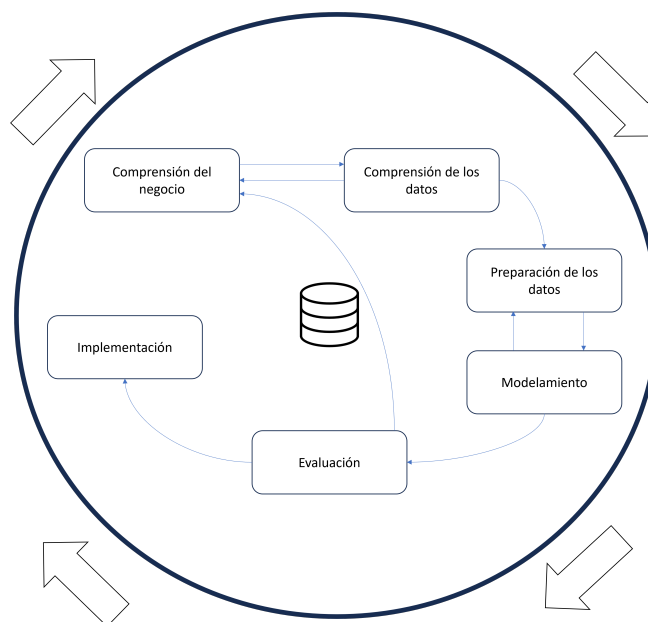


Figura 3: Representación de las etapas de CRISP-DM.
Fuente: Elaboración propia.

Como se observa en la Figura 3, la metodología CRISP-DM se caracteriza por su enfoque iterativo, lo que permite retornar a etapas anteriores cuando sea necesario. Esta flexibilidad la distingue frente a otras metodologías más lineales.

Por último, *Catalyst*, también conocida como *P3TQ* (*Product, Place, Price, Time, Quantity*), se estructura en dos modelos principales: el Modelo de Negocio (MII) y el Modelo de Exploración de Información (MIII) [Moine *et al.*, 2011]. El MII se enfoca en identificar problemas u

oportunidades empresariales y los requisitos específicos de la organización, proponiendo acciones según el contexto inicial. Para casos donde el problema no está claro, se recomienda explorar las relaciones *P3TQ* clave en la cadena de valor. Por otro lado, el MIII guía la construcción y ejecución de modelos de minería de datos basados en el MII. Esta metodología, centrada en la cadena de valor organizacional, se ha difundido como *P3TQ*, aunque no sea su nombre original, presentando una secuencia flexible de pasos interactivos para adaptarse a diversos caminos y evaluar continuamente los resultados antes de avanzar al siguiente paso.

2.2. Web Scraping

El *web scraping* es un conjunto de prácticas para extraer datos de la web. Antes de la llegada de las APIs (*Application Programming Interface*), el *web scraping* era la única forma para los programas de obtener información de internet. Las APIs, interfaces especiales para la comunicación entre aplicaciones y servidores, han revolucionado este proceso al ofrecer datos organizados, aunque no todas son gratuitas ni proporcionan amplia información [Khder, 2021].

Algunas técnicas comunes utilizadas en el *web scraping* incluyen el uso de bibliotecas como *curl* y *wget* para enviar solicitudes HTTP GET a la dirección de destino y obtener la página HTML como respuesta. Luego, en la etapa de extracción, se utilizan expresiones regulares, bibliotecas de análisis HTML y consultas XPath para extraer los datos importantes [Khder, 2021].

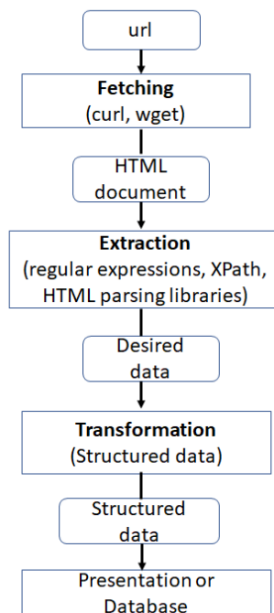


Figura 4: Proceso de Web Scraping.
Fuente: [Persson, 2019].

Para obtener datos desde páginas web de tipo dinámico, es posible utilizar herramientas que permitan simular la interacción de un usuario con la interfaz, lo que hace necesario recurrir a tecnologías específicas para la extracción de dicha información. Una de las herramientas más utilizadas para este propósito es Selenium, un entorno de automatización de pruebas web que permite ejecutar acciones simulando la interacción de un usuario real con los distintos elementos de una página. Esta herramienta ofrece compatibilidad con diversos lenguajes de programación, tales como Python, JavaScript, Java, entre otros, lo que facilita su integración en distintos entornos de desarrollo.

El *web scraping* enfocado en redes sociales permite extraer datos de manera automatizada de diferentes plataformas y aplicaciones de redes sociales. Esta práctica se ha vuelto cada vez más popular debido a la gran cantidad de información valiosa que se encuentra disponible en estas plataformas. Entre los datos que se pueden obtener se incluyen información demográfica, preferencias de los usuarios, opiniones, comentarios, publicaciones, contenido multimedia, entre otros.

Existen diversas herramientas y técnicas que facilitan el proceso de *web scraping* en redes sociales. Algunas plataformas incluso ofrecen APIs que permiten acceder a sus datos y extraerlos de forma más sencilla y estructurada. [Dewi y Chandra, 2019] realizó un trabajo de *web scraping* utilizando las APIs oficiales ofrecidas por Facebook (actualmente llamado Meta), para Facebook e Instagram y la API oficial de Twitter (actualmente llamado X), junto al uso de expresiones regulares.

Y combinar dicha extracción de información de redes sociales junto a distintas técnicas, como por ejemplo *machine learning*, puede llevar a descubrimientos poderosos, como el que se presenta en el trabajo realizado por [Li *et al.*, 2019] donde junto a la extracción de datos de Instagram y un modelo de aprendizaje profundo, lograron una herramienta para detectar vendedores de droga en la red social.

También es importante tener en cuenta que el *web scraping* en redes sociales debe ser realizado con ética y respetando las políticas y términos de uso establecidos por cada plataforma. Algunas redes sociales tienen restricciones o limitaciones en cuanto al acceso y extracción masiva de datos, por lo que es necesario asegurarse de cumplir con todas las normativas legales y éticas al realizar esta práctica.

2.3. Visualizaciones de datos

La visualización de datos es el proceso mediante el cual se representan datos e información mediante elementos visuales como gráficos, mapas, diagramas y otras formas de presentación visual. Su propósito principal es facilitar la comprensión de grandes volúmenes de información, permitiendo identificar con mayor claridad patrones, tendencias, relaciones o anomalías que podrían pasar desapercibidas en formatos textuales o numéricos [Sadiku *et al.*, 2016].

En el contexto contemporáneo, caracterizado por el auge del big data, la visualización de datos ha adquirido un papel central. La disponibilidad de enormes cantidades de datos provenientes de diversas fuentes —como sitios web, sensores, dispositivos inteligentes y plataformas de redes sociales— ha generado la necesidad de herramientas que permitan interpretar esta información de manera efectiva [Keim *et al.*, 2013]. En este escenario, la visualización de datos no solo facilita la exploración y análisis de la información, sino que también permite comunicar hallazgos de forma clara, intuitiva y atractiva.

Más allá de su función técnica, la visualización también cumple un rol narrativo: al estructurar y presentar los datos adecuadamente, se convierte en un medio para contar historias (*data storytelling*). Esto implica decisiones deliberadas sobre qué datos mostrar, cómo agruparlos y qué elementos destacar para guiar al espectador hacia una interpretación significativa [Dykes, 2016]. En este proceso, es fundamental encontrar un equilibrio entre la simplicidad visual y la fidelidad a la información original. Una visualización demasiado simplificada puede distorsionar la interpretación o eliminar detalles relevantes, mientras que una demasiado compleja puede dificultar su comprensión.

2.4. Periodismo de datos

El periodismo de datos es una disciplina que fusiona la narración periodística con el análisis de datos para descubrir, contar y comunicar historias [Molina, 2021]. Utiliza técnicas de recolección, limpieza, análisis y visualización de información para revelar patrones, tendencias o relaciones significativas en conjuntos de datos complejos. Esta práctica permite transformar datos en narrativas comprensibles, brindando una nueva dimensión a la forma en que se informa y se comprenden los acontecimientos en la sociedad actual.

En Chile destaca el trabajo de Ciudadanía Inteligente: generan visualizaciones de datos interactivas y atractivas sobre temas contingentes para la sociedad chilena, promoviendo la toma de decisiones informadas de una manera sencilla para el público.

Otro trabajo de periodismo de datos destacado en Chile es el realizado por DECIDECHILE, una plataforma de visualización y análisis de datos enfocada en temas públicos. Esta plataforma provee una gran variedad de visualizaciones sobre temas públicos en el país, inclusive con registros históricos como las elecciones presidenciales de 1989 hasta la conformación del Consejo Constitucional cuyas votaciones fueron el 7 de mayo de 2023.



Figura 5: Visualización sobre la conformación del Consejo Constitucional electo en mayo de 2023.

Fuente: [DECIDECHILE, 2023].

2.5. Análisis de Sentimientos

El análisis de sentimientos, también conocido como minería de opiniones, es un campo de estudio dentro del Procesamiento del Lenguaje Natural (PLN) que se enfoca en identificar, extraer y cuantificar estados afectivos y opiniones subjetivas a partir de datos textuales. Su objetivo principal es determinar la polaridad de un texto, es decir, clasificarlo como positivo, negativo o neutro. En el contexto de esta memoria, el análisis de sentimientos es una herramienta fundamental, ya que permite transformar el concepto abstracto de "percepción" sobre la salud mental en una métrica cuantificable, ofreciendo una visión directa de la reacción del público ante las publicaciones en redes sociales.

La aplicación del análisis de sentimientos al contenido de redes sociales es particularmente relevante debido al vasto volumen de datos de opinión que estas plataformas generan. Los comentarios, publicaciones y reacciones de los usuarios son una fuente rica de información sobre sus actitudes y emociones. Sin embargo, analizar estos textos presenta desafíos únicos, como la informalidad del lenguaje, el uso de sarcasmo, ironía, abreviaturas, emojis y un contexto cultural específico, que deben ser abordados para una correcta interpretación [Gomez-Adorno *et al.*, 2024].

Para llevar a cabo el análisis de sentimientos, se emplean diversas técnicas y algoritmos de *machine learning*. Estas se pueden agrupar en tres categorías principales:

1. Enfoques basados en léxicos: Utilizan diccionarios o "léxicos" de palabras a las que se les ha asignado previamente una puntuación de polaridad (positiva o negativa) e intensidad. El sentimiento general de un texto se calcula agregando las puntuaciones de las palabras que contiene. Aunque son sencillos de implementar, su eficacia puede ser limitada al no capturar el contexto o la negación [Gomez-Adorno *et al.*, 2024].
2. Aprendizaje supervisado: Este enfoque utiliza algoritmos de clasificación como Naive Bayes, Máquinas de Soporte Vectorial (SVM) o Regresión Logística. Estos modelos se

entrenan con un conjunto de datos previamente etiquetado por humanos (por ejemplo, tweets clasificados como positivos, negativos o neutros). Una vez entrenados, pueden predecir la polaridad de nuevos textos.

3. Redes Neuronales y Aprendizaje Profundo (*Deep Learning*): Modelos más avanzados, como las Redes Neuronales Recurrentes (RNN), *Long Short-Term Memory (LSTM)* y, más recientemente, los modelos basados en Transformers (ej. BERT), han demostrado una eficacia superior [Hossain *et al.*, 2025]. Estos modelos son capaces de comprender el contexto, la semántica y las relaciones complejas entre las palabras en una oración, lo que los hace especialmente poderosos para interpretar matices como el sarcasmo.

La incorporación del análisis de sentimientos es crucial para alcanzar los objetivos de esta memoria. Permite ir más allá de un simple conteo de publicaciones y analizar la naturaleza de la conversación pública sobre salud mental en Chile. Al clasificar la polaridad de las reacciones, es posible identificar temas que generan mayor preocupación, apoyo o estigma, y cómo esta percepción evoluciona en el tiempo.

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

En este capítulo se presenta detalladamente la solución desarrollada para comprender las percepciones sobre salud mental en una red social a definir. Aquí se describen las decisiones metodológicas adoptadas, los procesos técnicos ejecutados, y los resultados alcanzados mediante técnicas de extracción de datos, análisis de sentimientos y visualización de datos.

3.1. Selección de la metodología de trabajo

La metodología seleccionada para guiar el proceso de investigación fue el modelo CRISP-DM, desarrollado originalmente por IBM, debido a su amplio reconocimiento y aplicabilidad en diversos sectores. Este modelo estructura claramente las etapas necesarias para garantizar la calidad y relevancia del estudio realizado, y comprende seis fases fundamentales: **Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación, Despliegue.**

La elección del modelo CRISP-DM proporciona una estructura metodológica para el desarrollo de la investigación, lo que contribuye a organizar el proceso de análisis de datos y facilita la documentación de los procedimientos realizados; Además, destaca por su naturaleza iterativa. Si bien su versión tradicional no contempla explícitamente un proceso de periodismo de datos, este puede integrarse de manera complementaria dentro de sus etapas, adaptando el modelo a las necesidades específicas del estudio.

3.2. Comprensión del negocio

La primera etapa del proceso CRISP-DM consiste en comprender en profundidad los objetivos y requerimientos del proyecto desde una perspectiva de negocio, para luego traducirlos en metas específicas de minería de datos. En el contexto de esta investigación, el “negocio” se entiende como la necesidad de analizar la percepción social sobre la salud mental en Chile a través de las reacciones de usuarios en una red social, utilizando técnicas de *machine learning* y periodismo de datos.

3.2.1. Determinación de los objetivos del negocio

El objetivo general de esta memoria es desarrollar un análisis de las reacciones de los usuarios ante publicaciones relacionadas con la salud mental en Chile, aplicando técnicas de *ma-*

chine learning y periodismo de datos, con el propósito de aportar a una mejor comprensión de la percepción social sobre el tema.

A partir de este objetivo general, se establecen los objetivos específicos indicados en la Sección 1.2.

Estos objetivos definen claramente el alcance del proyecto: entender el contexto actual, recolectar datos relevantes, procesarlos mediante herramientas de análisis de sentimientos, y presentar los hallazgos de manera comprensible y accesible.

3.2.2. Evaluación de la situación

La evaluación de la situación considera diversos aspectos relevantes para el desarrollo de esta investigación. En primer lugar, se identificaron los recursos disponibles para llevar a cabo el análisis. Entre estos se encuentra el acceso a datos públicos de redes sociales, como publicaciones provenientes de plataformas como Instagram o Twitter. Asimismo, se cuenta con herramientas de procesamiento de lenguaje natural y modelos de análisis de sentimientos preentrenados, lo que facilita el tratamiento inicial de los textos recopilados. Por otro lado, se dispone de plataformas de visualización de datos (como Power BI o Tableau) y diversas bibliotecas de visualización en Python, como Seaborn, Plotly y Matplotlib, que permiten representar de manera gráfica los resultados obtenidos.

No obstante, también se reconocen ciertas limitaciones que podrían afectar el desarrollo de este trabajo. Entre ellas se encuentran las restricciones de acceso a determinados datos debido a las políticas de privacidad propias de cada plataforma, así como la disponibilidad de herramientas gratuitas que permitan realizar *scraping* en la red social seleccionada. Además, se identifican posibles sesgos en los datos recopilados, como la sobrerrepresentación de usuarios más activos o de determinados grupos etarios. A esto se suma la consideración de que los modelos de análisis de sentimientos, al estar entrenados en *corpus* de uso general, podrían no captar de manera adecuada las particularidades lingüísticas y culturales del contexto chileno en relación con la salud mental.

Finalmente, se detectaron algunos riesgos asociados al proceso de recopilación y análisis de datos. Dentro de estos se encuentra la calidad variable de los datos obtenidos, que podría incluir errores, duplicados o textos irrelevantes. Asimismo, se anticipan dificultades en la interpretación automática de ciertos recursos comunicativos propios del lenguaje en redes sociales, como la ironía o el sarcasmo, lo cual podría afectar la precisión del análisis de sentimientos.

3.3. Comprensión de los datos

La fase de Comprensión de los Datos tiene como objetivo explorar y familiarizarse con los datos que serán utilizados en el análisis. Esta etapa resulta fundamental para garantizar que los datos recopilados sean relevantes, representativos y adecuados para alcanzar los objetivos del proyecto.

En el contexto de esta investigación, esta fase contempla varias actividades esenciales: la selección de la red social desde donde se obtendrán las publicaciones relacionadas con salud mental en Chile, el establecimiento de criterios para la elección de las publicaciones a analizar, el diseño y ejecución del proceso de extracción de los datos, y finalmente, la descripción detallada de las características principales del conjunto de datos obtenido.

En esta etapa se incluye también la contextualización de los datos, la cual se desarrolla a continuación.

3.3.1. Contextualización: Estadísticas sobre salud mental en Chile

Para contextualizar adecuadamente el estado actual de la salud mental en Chile, se consideraron múltiples fuentes relevantes que proporcionan estadísticas y métricas significativas de los últimos años. Estas fuentes describen la realidad del país en torno a este tema fundamental. La elaboración de esta sección se basó en el análisis de diversas fuentes de datos respaldadas sobre salud mental.

Entre las fuentes seleccionadas más destacadas se encuentran:

- La Encuesta Nacional de Salud, Sexualidad y Género (ENSSEX) 2022-2023²: Un estudio desarrollado por el Departamento de Epidemiología del Ministerio de Salud de Chile, en colaboración con la Universidad de Chile, la Universidad Alberto Hurtado y el Institut National d'Études Démographiques de París. Su objetivo es monitorear el estado de salud de la población y aportar evidencia para el diseño, implementación y evaluación de políticas públicas. Este estudio se enfoca en personas de 18 años y más residentes en Chile, y cuenta con representatividad a nivel nacional, regional, por tramos de edad y sexo.
- El Atlas de Salud Mental³: Publicado por la Organización Mundial de la Salud (OMS)⁴ cada tres años, es una recopilación de información proporcionada por distintos países

²ENCUESTA NACIONAL DE SALUD, SEXUALIDAD Y GÉNERO (ENSSEX) 2022-2023 - <https://epi.minsal.cl/ensex-que-es-la-encuesta/>

³World Health Organization. (2021). *Mental Health Atlas 2020*. <https://www.who.int/publications/i/item/9789240036703>

⁴Organización Mundial de la Salud (OMS). Quiénes somos. Recuperado el 3 de mayo de 2025, de <https://www.who.int/es/about/who-we-are>

sobre políticas, legislación, financiamiento, recursos humanos, disponibilidad y utilización de servicios, y sistemas de recopilación de datos en salud mental. La OMS es una entidad internacional que coordina la respuesta global frente a emergencias sanitarias, promueve el bienestar, previene enfermedades y busca ampliar el acceso equitativo a servicios de salud. Integrada por más de 8.000 profesionales especializados en salud pública, incluyendo médicos, epidemiólogos, científicos y administradores, la OMS trabaja proporcionando evidencia científica sólida a países, comunidades y colaboradores, con el objetivo de garantizar que todas las personas tengan igualdad de oportunidades para llevar una vida segura y saludable.

- El Termómetro de la Salud Mental en Chile Achs-UC ⁵: Estudio longitudinal ⁶ realizado desde 2020, que analiza el estado de la salud mental en la población urbana chilena mayor de 18 años. Este estudio utiliza una muestra aleatoria de aproximadamente 3.000 individuos, representativa a nivel nacional, seleccionada a partir de hogares previamente entrevistados presencialmente. La información se recoge mediante entrevistas telefónicas periódicas, lo que permite monitorear los cambios y tendencias en salud mental a lo largo del tiempo. La Asociación Chilena de Seguridad (ACHS) ⁷ es una corporación privada sin fines de lucro que entrega protección y atención en seguridad laboral a más de tres millones de trabajadores afiliados y a más de 73.000 entidades empleadoras en Chile. Destaca por registrar la menor tasa promedio de accidentabilidad laboral, consolidándose como la mutualidad más relevante del país en su ámbito. La Pontificia Universidad Católica de Chile (PUC) es una institución líder en educación superior, reconocida por su excelencia académica e investigación de alto impacto.
- El Informe Anual de de Seguridad y Salud en el Trabajo: Un informe elaborado por la Superintendencia de Seguridad Social (SUSESO) que proporciona datos sobre denuncias por accidentes laborales y enfermedades profesionales en Chile, incluyendo el número de casos, tasas de accidentabilidad y desagregaciones detalladas, además de información sobre las prestaciones del seguro contemplado en la Ley N°16.744. La SUSESO es un organismo autónomo del Estado encargado de fiscalizar el cumplimiento de la normativa en materia de seguridad social y velar por la protección de los derechos de los trabajadores, pensionados y sus familias.
- El Observatorio de Salud Mental Chile: Organización multidisciplinaria dedicada a investigar, evaluar e intervenir en problemáticas psicosociales, promoviendo el bienestar y calidad de vida de las personas mediante acciones concretas basadas en evidencia.

⁵Asociación Chilena de Seguridad (ACHS). (2024, 7 de junio). Novena ronda del Termómetro de la Salud Mental ACHS-UC [Noticia]. Recuperado el 3 de mayo de 2025, de <https://www.achs.cl/centro-de-noticias/noticia/2024/novena-ronda-del-term%C3%B3metro-de-la-salud-mental-achs-uc>

⁶Un estudio longitudinal es aquel que investiga o recolecta información sobre un mismo grupo de individuos o una misma muestra en distintos momentos del tiempo. Este tipo de estudio permite observar cómo cambian o evolucionan ciertas características, comportamientos o fenómenos en la población estudiada.

⁷Asociación Chilena de Seguridad (ACHS). (s.f.). *Marco Regulatorio - Ley 16.744*. <https://www.achs.cl/nosotros/achs-corporativo/marco-regulatorio-ley-16744>

El impacto de los problemas de Salud Mental

Según el Mental Health Atlas 2020 Country Profile: Chile [World Health Organization, 2022a], Chile presentó una tasa de Años de Vida Ajustados por Discapacidad (DALY, por sus siglas en inglés) de 1.994,2 por cada 100.000 habitantes, lo que evidencia un impacto significativo de los trastornos mentales en la calidad de vida de la población. Sin embargo, al contrastar esta cifra con la correspondiente a los años 2017 [World Health Organization, 2017] y 2014 [World Health Organization, 2014], que se situaba en aproximadamente 3.109, se observa una disminución en términos cuantitativos. No obstante, esta variación no permite afirmar con certeza una mejora en el estado de salud mental del país, dado que podrían incidir otros factores, tales como modificaciones en las metodologías de medición, entre otras posibles causas.

En 2022, el 67% de las enfermedades reportadas por profesionales en Chile y de origen laboral estuvieron relacionadas con salud mental, aumentando significativamente respecto al año anterior (52%) [Superintendencia de Seguridad Social, 2023], en el año 2024 esta cifra fue de 72% [Superintendencia de Seguridad Social, 2025]. Como medida de prevención, el Cuestionario para la Evaluación de Ambientes de Trabajo Saludables en Salud Mental (CEAL-SM) se ha vuelto obligatorio para todas las empresas chilenas [Superintendencia de Seguridad Social, 2023], destacando que el agotamiento laboral (burnout) y la violencia y acoso laboral están directamente relacionados con síntomas depresivos [Asociación Chilena de Seguridad, 2025].

Prevalencia de enfermedades mentales

Como antecedente se menciona un estudio realizado en 2004 [Vicente *et al.*, 2004] el cual reveló que la prevalencia de trastornos psiquiátricos en Chile era del 19.7% en el sexto mes desde la encuesta y del 16.7% en el primer mes. Los trastornos más comunes identificados a los seis meses fueron la fobia simple, la fobia social, la agorafobia, el trastorno depresivo mayor y la dependencia del alcohol. Otro aspecto muy interesante mencionado en este estudio es la conclusión de que en Chile la tasa de consulta a servicios de salud mental es alta en personas con trastornos psiquiátricos comórbidos⁸. Si bien este estudio proporciona una referencia inicial, es fundamental considerar datos más recientes para comprender la situación actual.

La encuesta Termómetro de Salud Mental [Bravo *et al.*, 2024], en su octava ronda, indica que un 13,1% presentaba síntomas depresivos moderados o severos, cifra que se ha mantenido constante desde las últimas versiones de la encuesta. De quienes tuvieron síntomas, solo un 40,7% logró efectivamente consultar con un profesional. Los resultados también indican que la ansiedad continúa siendo el trastorno más prevalente: en esta ronda, el 24.8% de los encuestados manifestó síntomas de ansiedad, cifra ligeramente superior al 22.3% previo,

⁸La comorbilidad se refiere a la presencia simultánea de dos o más trastornos o enfermedades en una misma persona. En el contexto de la salud mental, esto significa que una persona puede tener, por ejemplo, depresión y ansiedad generalizada al mismo tiempo, o un trastorno por consumo de sustancias junto con un trastorno de pánico.

consolidando una tendencia estable desde mediados de 2021.

Los resultados de la Encuesta Nacional de Salud, Sexualidad y Género [Ministerio de Salud, 2025] revelan que las mujeres presentan mayores indicadores de malestar psicológico en comparación con los hombres. Mientras un 60,9% de los hombres afirmó no tener síntomas, esta cifra disminuye al 46,1% en el caso de las mujeres. Por otro lado, el 29,7% de los hombres y el 36,6% de las mujeres reportó síntomas leves; un 7% de los hombres y un 12,4% de las mujeres presentó síntomas moderados; y un 1,9% de los hombres, frente a un 4,5% de las mujeres, manifestó síntomas severos. Todas estas diferencias resultaron ser estadísticamente significativas, evidenciando una mayor carga de malestar psicológico en la población femenina.

Al considerar la variable de orientación sexual, se observa que un 54,2% de las personas heterosexuales declaró no presentar síntomas de malestar psicológico, mientras que esta cifra disminuye al 37% entre personas gays, lesbianas, bisexuales y otras identidades no heterosexuales. En este último grupo, se evidencian diferencias estadísticamente significativas en la presencia de síntomas moderados, con un 15,6%, cifra considerablemente mayor en comparación con el 9,5% registrado en la población heterosexual [Ministerio de Salud, 2023].

Al comparar las estadísticas recientes con las del estudio citado al inicio de esta sección, se evidencia un cambio significativo en el panorama de la salud mental en Chile. Mientras que en años anteriores los trastornos más frecuentes estaban relacionados con fobias y adicciones, en la actualidad predominan condiciones como la ansiedad y la depresión en sus distintas variantes.

Percepción y estigma

Sobre la percepción social, el Observatorio de Salud Mental reportó en octubre de 2024 que un 87.19% considera baja la efectividad de la Ley Karin contra el acoso laboral [Observatorio de Salud Mental Chile, 2024]. Asimismo, menos de la mitad de quienes padecen ansiedad o depresión reconocen la necesidad de atención especializada, reflejando la importancia de mejorar la educación e información sobre salud mental [Bravo *et al.*, 2024].

La creciente percepción ciudadana sobre la gravedad de los problemas mentales también es notoria: la identificación de la salud mental como un problema principal de salud pasó del 26% en 2018 al 69% en 2024, influenciada probablemente por la pandemia de COVID-19, que incrementó significativamente los trastornos mentales comunes [Ipsos, 2024].

Un 79.7% de la población joven considera que se le da poca importancia a los problemas de salud mental en Chile [Instituto Nacional de la Juventud, 2019].

Un estudio realizado en 2019 entre trabajadores chilenos encontró que un 85% cree que la mayoría de los empleadores no contratarían a una persona que ha sido hospitalizada por una enfermedad mental. Además, el 80% siente pena por las personas con enfermedades mentales graves, y el 78% cree que la mayoría de los hombres no se casarían con una mujer

con una enfermedad mental [Solis-Soto *et al.*, 2019].

Una encuesta realizada por INJUV en 2019 entre jóvenes de 15 a 29 años reveló que un 87.4% está de acuerdo en que las personas con problemas de salud mental evitan decirlo por temor a ser discriminadas. Además, un 78.0% de los jóvenes cree que en la sociedad chilena está mal visto ir al psicólogo o psiquiatra por problemas de salud mental [Instituto Nacional de la Juventud, 2019].

El estigma asociado a la salud mental también se manifiesta en las coberturas de salud. Según el Departamento de Estudios y Desarrollo, persiste una brecha considerable entre la cobertura financiera efectiva de las prestaciones en salud mental en comparación con las correspondientes a la salud física, así como en la resolución de licencias médicas por trastornos mentales frente a otros grupos diagnósticos [Superintendencia de Salud, 2023].

3.3.2. Selección de la plataforma para el análisis

Elegir una plataforma adecuada para realizar este análisis fue un paso crítico para asegurar resultados relevantes y de alta calidad. Inicialmente se evaluaron Instagram, YouTube y TikTok, plataformas que destacan por sus niveles de interacción, diversidad de formatos y alta popularidad en la sociedad chilena actual. Se priorizaron formatos breves y dinámicos, especialmente por su capacidad de atraer interacciones significativas en torno a temas sensibles como la salud mental.

A continuación, se muestra la definición de los formatos más destacados.

- **Reels (Instagram):** Videos breves diseñados para captar la atención rápidamente, permitiendo conectar creativamente con una audiencia diversa mediante contenidos frescos y dinámicos [Instagram24].
- **Shorts (YouTube):** Segmentos cortos de hasta 60 segundos que estimulan interacciones espontáneas y facilitan la creación y difusión rápida de tendencias [YouTube24].

Aunque TikTok posee una gran base de usuarios activos, fue descartada debido a las limitaciones técnicas y la falta de soluciones gratuitas para extraer información específica. YouTube también fue descartado, principalmente porque su API oficial no permite la extracción efectiva del contenido tipo Shorts. Instagram emergió finalmente como la opción más viable, dado su volumen masivo de contenido diario, alta interacción de usuarios y uso extendido de *hashtags*, factores clave para facilitar el filtrado y análisis de datos. Aunque Instagram carece de una solución oficial completa para la extracción de datos, existía suficiente documentación técnica que permitía desarrollar una solución específica adaptada a los objetivos de esta investigación. Esto implicó extender el trabajo planificado en un principio.

3.3.3. Identificación de *hashtags* relevantes

La selección rigurosa de *hashtags* adecuados fue crucial para obtener un dataset relevante. Se exploraron inicialmente *hashtags* específicos relacionados directamente con trastornos mentales (por ejemplo, DepresionChile y AnsiedadChile), pero debido al bajo volumen de publicaciones se decidió ampliar el alcance hacia términos más generales. Finalmente, tras un análisis exhaustivo, los *hashtags* SaludMentalChile y PsicologiaChile demostraron ser los más apropiados, capturando una mayor cantidad de información relevante.

3.3.4. Extracción de datos: desafíos y soluciones

Uno de los principales desafíos de esta memoria fue la extracción efectiva de datos, debido a recientes modificaciones técnicas en Instagram que limitaron el acceso a la información a través de Instaloader, la herramienta de código abierto más utilizada para realizar scraping en esta red social. Estas actualizaciones, muchas de ellas no documentadas oficialmente, afectaron especialmente la funcionalidad de búsqueda de publicaciones por *hashtag*, un componente esencial para el desarrollo de este trabajo. Ante esta situación, fue necesario explorar alternativas que permitieran continuar con la recopilación de datos de manera adecuada.

Finalmente, tras varios intentos, la solución implicó el desarrollo de un enfoque híbrido que combinó una versión de prueba de una API comercial con un programa propio especialmente diseñado para recolectar los comentarios asociados a cada publicación. La API seleccionada, después de probar distintas opciones, fue Instagram *hashtag* Scraper de Apify⁹. Este enfoque permitió superar las barreras inicialmente planteadas, generando dos primeros dataset (uno para el *hashtag* SaludMentalChile y otro para PsicologiaChile) para análisis posteriores.

Es importante mencionar otras limitaciones no previstas que surgieron al recolectar los datos:

- Instagram eliminó la opción de filtrar publicaciones por fecha al realizar búsquedas por *hashtag*, razón por la cual los datos obtenidos presentan una distribución temporal desconocida inicialmente. Para abordar este problema se incluye en el análisis exploratorio de los datos, un estudio específico de las fechas de las publicaciones obtenidas.
- La versión de prueba de la API comercial utilizada no permitía controlar el punto de continuación de las búsquedas, es decir, no era posible reanudar una consulta desde una página de resultados determinada (por ejemplo, desde la página 9 tras finalizar la página 8). Como consecuencia, cada nueva ejecución de la herramienta iniciaba nuevamente desde el principio, dificultando la incorporación de datos nuevos en iteraciones posteriores. Esta limitación, junto con las restricciones propias de la prueba gratuita de

⁹<https://apify.com/apify/instagram-hashtag-scraper>

la API, redujo la cantidad de datos que se pudieron recolectar. No obstante, se logró obtener un volumen de datos suficientemente representativo numéricamente para realizar un análisis sólido.

3.3.5. Descripción de los datos

En esta sección se describen los datos y sus estructuras, utilizando tablas y visualizaciones para facilitar la comprensión.

En las siguientes visualizaciones se detalla la distribución temporal que tienen los datos recolectados, para cada *hashtag* en particular y para el conjunto en total.

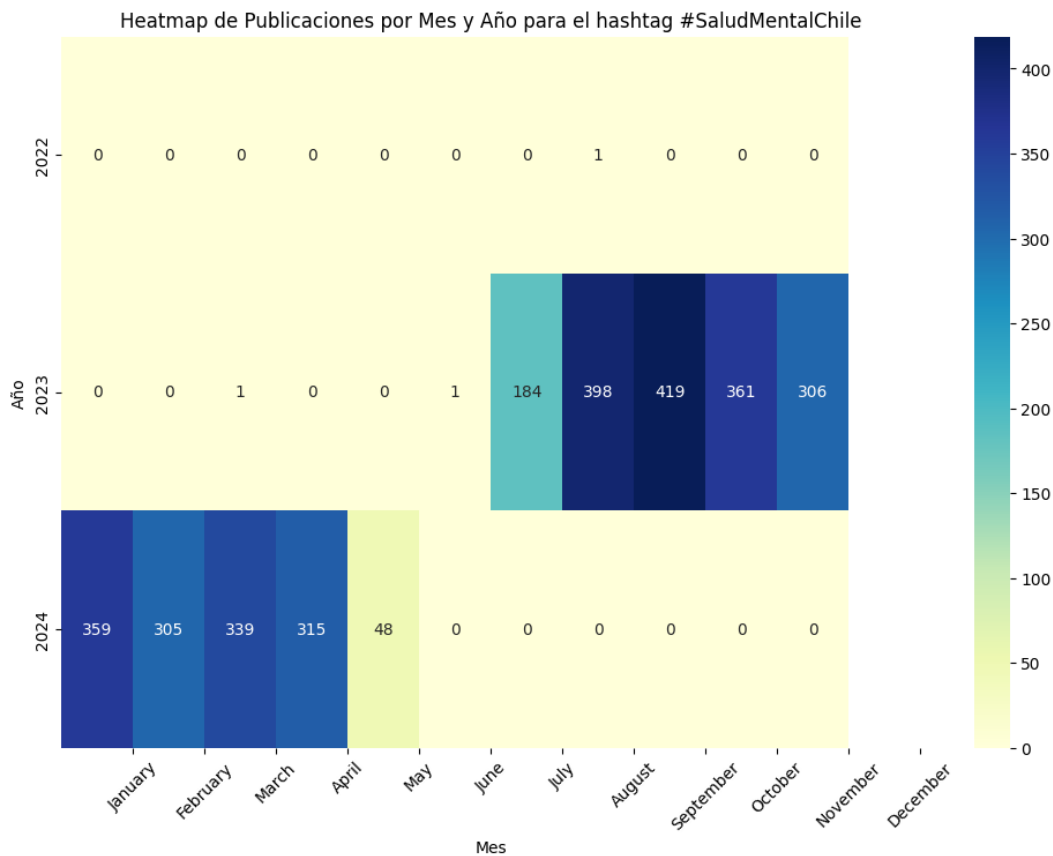


Figura 6: Mapa de calor de las publicaciones por mes y año para el *hashtag* SaludMentalChile.
 Fuente: Elaboración propia.

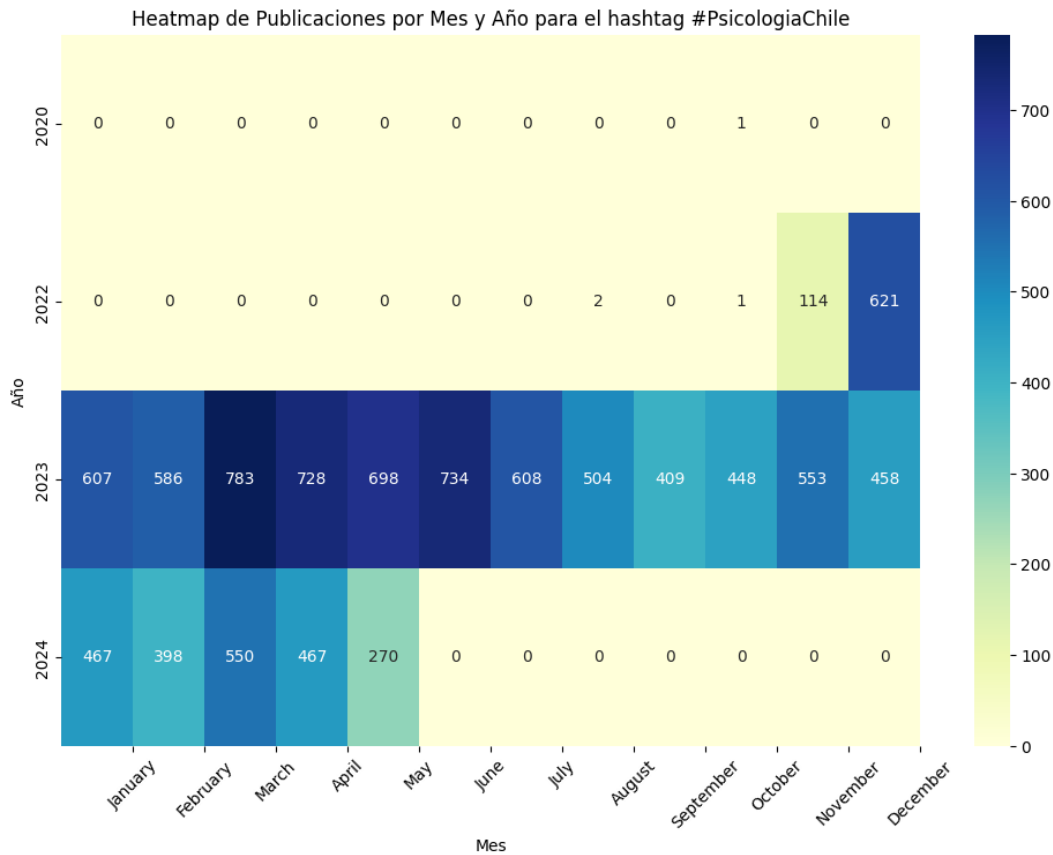


Figura 7: Mapa de calor de las publicaciones por mes y año para el *hashtag* PsicologiaChile.
 Fuente: Elaboración propia.

El análisis de las Figuras 6 y 7, correspondientes a los mapas de calor de publicaciones para los *hashtags* SaludMentalChile y PsicologiaChile respectivamente, permite identificar patrones temporales diferenciados en la actividad de usuarios de Instagram en torno a estas temáticas. Es importante señalar que ambas visualizaciones se construyeron a partir de muestras aleatorias de publicaciones obtenidas mediante *web scraping*, por lo que los valores registrados en cada celda no reflejan el total de publicaciones reales, sino una aproximación a través de los datos extraídos. En este sentido, se observan variaciones en la cantidad de publicaciones obtenidas por mes y año, sin que exista una correspondencia proporcional con la actividad total en la plataforma.

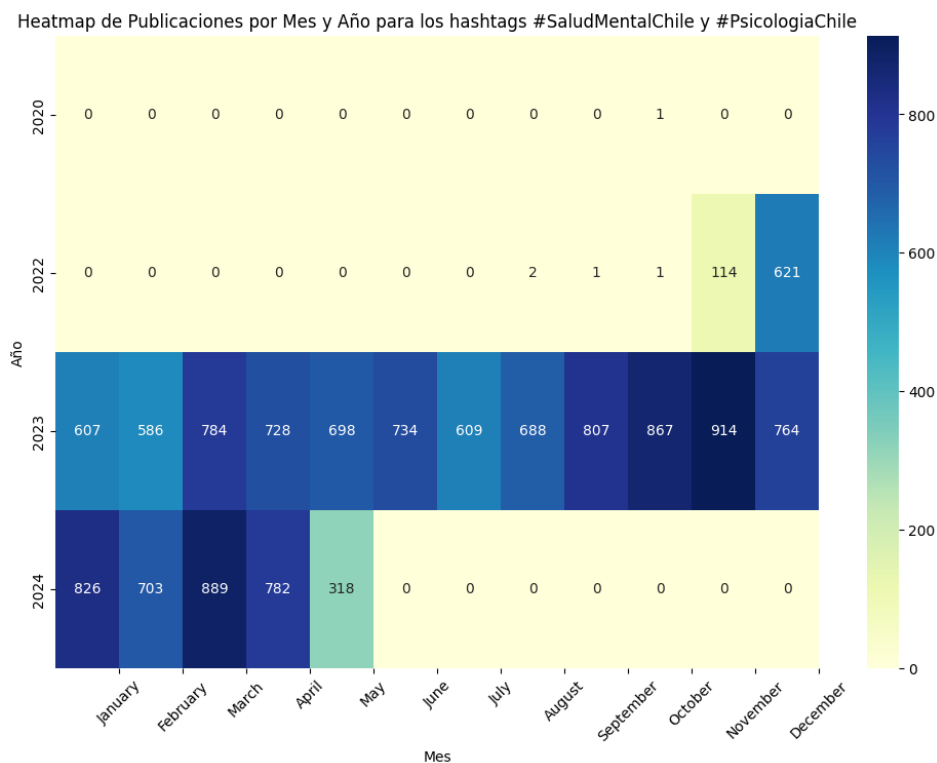


Figura 8: Mapa de calor de las publicaciones por mes y año para los *hashtags* SaludMentalChile y PsicologiaChile.

Fuente: Elaboración propia.

En particular, para el *hashtag* SaludMentalChile se aprecia un incremento de publicaciones en la muestra a partir de mediados de 2023, con un máximo en agosto y una concentración de publicaciones en los primeros meses de 2024. Por su parte, el *hashtag* PsicologiaChile evidencia una mayor y más sostenida actividad en la muestra desde comienzos de 2023, con valores elevados durante todo el primer semestre y un descenso gradual hacia fin de año. Además, se observan valores notoriamente altos en diciembre de 2022 y diciembre de 2023, lo que sugiere posibles picos de interés en esos períodos, aunque debido al carácter aleatorio y limitado de la recolección, estos comportamientos deben interpretarse con cautela. En síntesis, si bien ambas muestras sugieren una mayor actividad durante el año 2023, la temporalidad y distribución de las publicaciones asociadas a cada *hashtag* presentan diferencias significativas. Cabe señalar que, a la fecha de la recolección de datos (abril de 2024), se conoce que el volumen total de publicaciones asociadas al *hashtag* PsicologiaChile era de aproximadamente 15.000, mientras que para SaludMentalChile alcanzaba cerca de 40.000 publicaciones. Esta diferencia en el volumen total disponible se refleja en la mayor densidad de datos observada en la muestra extraída para este último.

Finalmente, la Figura 8 integra ambos conjuntos de datos con el objetivo de analizar la distribución total de publicaciones, observándose que el mayor volumen se concentra en noviembre de 2023.

Estructura de los datos

A continuación, se presentan tablas que permiten visualizar y describir la estructura de los datos utilizados. Estas tablas contienen una representación de las variables extraídas y organizadas tras el proceso de recolección mediante técnicas de *web scraping* y uso de API, proporcionando así una visión general del formato y contenido de los conjuntos de datos.

Tabla 1: Descripción de los atributos de un *post* obtenido desde la API utilizada.
 Fuente: Elaboración propia.

Atributo	Tipo de dato	Descripción
inputUrl	String	URL de búsqueda de Instagram basada en el <i>hashtag</i> .
id	String	ID única del <i>post</i> .
type	String	Tipo de contenido (por ejemplo, <i>Image</i> o <i>Sidecar</i> (carrusel)).
shortCode	String	Código corto para construir la URL del <i>post</i> .
caption	String	Texto del <i>post</i> (a veces vacío).
hashtags	Arreglo de String	<i>hashtags</i> incluidos en el <i>post</i> .
mentions	Arreglo de String	Usuarios mencionados en el <i>post</i> .
url	String	Enlace directo al <i>post</i> en Instagram.
commentsCount	Entero	Número de comentarios.
firstComment	String	Primer comentario del <i>post</i> .
latestComments	Arreglo de String	Últimos comentarios.
dimensionsHeight	Entero	Altura de la imagen en píxeles.
dimensionsWidth	Entero	Ancho de la imagen en píxeles.
displayUrl	String	Enlace directo a la primera imagen del <i>post</i> .
images	Arreglo	Imágenes adicionales en el <i>post</i> (puede estar vacía).
alt	String	Texto alternativo o descripción de la imagen. Instagram agrega su posible interpretación de la imagen en caso de no estar.
likesCount	Entero	Cantidad de “me gusta” en el <i>post</i> .
timestamp	DateTime	Fecha y hora en UTC de publicación del <i>post</i> .
childPosts	Arreglo	Publicaciones asociadas (por ejemplo, en carruseles).
ownerFullName	String	Nombre completo del usuario que publicó.
ownerUsername	String	Nombre de usuario de Instagram.
ownerId	String	ID única del usuario.
isSponsored	Booleano	Indica si el <i>post</i> es patrocinado o no.
fullName	String	Otro campo de nombre (en este dataset toma siempre el valor <i>owner didnt open</i>).

Tabla 2: Descripción de los atributos de un comentario en la estructura de datos.
 Fuente: Elaboración propia.

Atributo	Tipo de dato	Descripción
pk	String	Identificador único del comentario.
user_id	String	ID del usuario que escribió el comentario.
type	Entero	Tipo de comentario (ej. 0 para normal, 2 para respuesta).
did_report_as_spam	Booleano	Indica si fue reportado como spam.
created_at, created_at_utc	Timestamp	Fecha de creación del comentario (formato UNIX).
content_type	String	Tipo de contenido, en este caso <i>comment</i> .
status	String	Estado del comentario (por ejemplo, "Active").
share_enabled	Booleano	Indica si se puede compartir el comentario.
is_ranked_comment	Booleano	Si es un comentario destacado.
media_id	String	ID del <i>post</i> al que pertenece el comentario.
comment_index	Entero	Posición del comentario en la lista.
text	String	Texto del comentario.
has_liked_comment, has_disliked_comment	Booleano	Si el usuario ha dado "me gusta" o "no me gusta".
comment_like_count	Entero	Cantidad de "me gusta" que tiene el comentario.
has_translation	Booleano	Indica si hay traducción disponible.
child_comment_count	Entero	Número de comentarios como respuesta.
preview_child_comments	Arreglo	Lista de algunos comentarios como respuesta (si existen).
user	Objeto	Información del usuario autor del comentario.
user.pk, user.username, user.full_name	String	ID, nombre de usuario y nombre completo del autor.
user.is_verified, user.is_private	Booleano	Si el usuario está verificado y/o tiene cuenta privada.
user.profile_pic_url	String	URL de la imagen de perfil del autor.
is_created_by_media_owner	Booleano	Si el comentario fue hecho por quien publicó el <i>post</i> .
replied_to_comment_id	String	ID del comentario al que se está respondiendo. Solo en comentarios que son respuesta a otro.

Publicaciones

Como ya se indicó, se utilizó una API comercial para obtener las publicaciones de Instagram que contenían el *hashtag* SaludMentalChile. La respuesta de esta API se encuentra en formato JSON, y cada objeto representa una publicación individual. Cada objeto contiene información estructurada sobre el *post*, como su identificador, tipo de contenido, texto, *hashtags*, usuario que publica, métricas de interacción (comentarios y “me gusta”), dimensiones de la imagen, y enlaces tanto al *post* como a la imagen. Esta estructura de datos se describe en la Tabla 1. El mismo proceso se aplicó para el *hashtag* PsicologíaChile.

Comentarios

Dado que para obtener los comentarios se desarrolló en Python un programa propio, se conoce completamente la estructura utilizada por la aplicación para mostrar esta información (a diferencia de la información obtenida por la API comercial). Dicha estructura se presenta originalmente en la Tabla 2, aunque algunos elementos fueron omitidos debido a que se desconoce por completo su significado, ya que Instagram no proporciona documentación oficial sobre sus estructuras de datos.

La estructura representada en las siguientes tablas corresponde al formato jerárquico con el que se almacenan los comentarios asociados a una publicación específica en el dataset. Cada publicación está identificada por un *post_id*, el cual contiene un conjunto de comentarios indexados por su identificador único (*pk*). Cada comentario incluye atributos como el texto del mensaje, la información del usuario que lo escribió, la fecha de publicación y si fue reportado como spam o creado por el autor del *post*. Además, algunos comentarios pueden contener un subgrupo de respuestas directas almacenadas en el objeto *childs*, el cual replica la misma estructura que el comentario principal.

Tabla 3: Descripción de la estructura de comentarios jerárquicos asociados a publicaciones.
 Fuente: Elaboración propia.

Atributo	Tipo de dato	Descripción
<i>post_id</i>	String	Identificador único de la publicación a la que pertenecen los comentarios.
<i>comments</i>	Objeto	Conjunto de comentarios asociados al <i>post</i> , indexados por su <i>pk</i> .

Hashtags

El *hashtag* es una cadena de palabras precedida por el signo numeral (#), que convierte el texto en un metadato fácilmente identificable por plataformas digitales, sistemas y usuarios. Su función principal es agrupar contenidos relacionados en redes sociales, facilitando su búsqueda y visibilidad.

Tabla 4: Descripción de los atributos de un comentario y sus posibles respuestas (estructura jerárquica).

Fuente: Elaboración propia.

Atributo	Tipo de dato	Descripción
pk	String	Identificador único del comentario.
text	String	Texto del comentario.
user.user_id	String	ID del usuario que realizó el comentario.
user.full_name	String	Nombre completo del autor del comentario.
did_report_as_spam	Booleano	Indica si el comentario fue reportado como spam.
created_at_utc	Timestamp	Fecha y hora de publicación del comentario en formato UNIX UTC.
is_created_by_media_owner	Booleano	Indica si el comentario fue hecho por el autor del post.
childs	Objeto	Comentarios hijos de este comentario, indexados por su pk. Tienen la misma estructura que un comentario original.

En la Tabla 5 se muestra la cantidad de veces que aparecía cada *hashtag* de los primeros 10 resultados obtenidos de mayor a menor frecuencia.

Tabla 5: Hashtags más frecuentes en el dataset obtenido.

Fuente: Elaboración propia.

Hashtag	Cantidad
saludmentalchile	2683
saludmental	1622
ansiedad	697
psicologia	605
terapia	590
terapiaonline	583
instachile	580
amorpropio	548
psicoterapia	468
ayuda	441

Dado que los *hashtags* pueden presentarse en distintos formatos como «Depresión», «Depresion» o «depresión», y también pueden contener emojis, se realizó un preprocesamiento para uniformarlos. Este procedimiento consistió en normalizar los textos mediante Unicode,

eliminando tildes y otros signos diacríticos, así como en remover los emojis. Adicionalmente, como parte del preprocesamiento, se convirtió todo el texto a minúsculas y se eliminaron otros signos de puntuación. Finalmente, se extrajeron los *hashtags* únicos, evitando repeticiones. Este proceso permitió obtener métricas más relevantes, como las presentadas en la siguiente tabla.

Tabla 6: Hashtags más frecuentes en publicaciones con el *hashtag* SaludMentalChile.
 Fuente: Elaboración propia.

Hashtag	Cantidad
saludmentalchile	3034
saludmental	1753
psicologia	895
ansiedad	708
terapia	631
amorpropio	610
terapiaonline	600
instachile	584
psicoterapia	514
psicologiachile	491

Tabla 7: Hashtags más frecuentes en el dataset del *hashtag* PsicologíaChile.

Hashtag	Cantidad
psicologiachile	10163
psicologia	5567
saludmental	3596
psicoterapia	2479
psicologiaclinica	2365
chile	1957
terapia	1813
terapiaocupacionalchile	1420
perrossocializados	1370
perrosguardianesdeniños	1370

En las Tablas 6 y 7 se puede observar que naturalmente los *hashtags* que aparecen más veces son aquellos por los que se realizó la búsqueda, lo cual es coherente.

También se identifican *hashtags* que, si bien aparecen con alta frecuencia, no tienen relación semántica directa con temas de salud mental. Por ejemplo, algunos *hashtags* corresponden a etiquetas de diseño gráfico, marketing o dinámicas de crecimiento en redes sociales. Estos fueron analizados y, en algunos casos, eliminados por no ser relevantes para evitar sesgos en el análisis posterior. No obstante, es importante aclarar que no necesariamente esto implica

que la publicación sea eliminada del *dataset*. La cantidad de *hashtags* totales entre ambos *datasets* es de 213.143.

Se observa que la mayoría de los *hashtags* aparecen una sola vez, lo cual indica una alta dispersión en el uso de estos. Esta información es útil para identificar qué proporción del conjunto de datos está compuesta por términos poco frecuentes.

Por último, la Tabla 8 muestra la cantidad de comentarios totales por *hashtag*.

Tabla 8: Cantidad de comentarios asociados a los *hashtags* analizados.
 Fuente: Elaboración propia.

Hashtag	Cantidad de comentarios
SaludMentalChile	4.921
PsicologiaChile	9.787

3.4. Preparación de los datos

En esta sección se describen los procesos realizados para preparar los datos con el fin de llevar a cabo el análisis. Una parte inicial de este procedimiento fue presentada en la sección anterior; aquí se continúa y profundiza en los pasos aplicados. Todas las tareas descritas fueron implementadas utilizando el lenguaje de programación Python.

Para el análisis de los comentarios, se construyó un nuevo conjunto de datos a partir de una selección de los datos originales, generando un archivo en formato CSV con la estructura mostrada en la Tabla 9. Los atributos incluidos mantienen el mismo significado detallado previamente en la sección de descripción de los datos, con la excepción de un nuevo atributo denominado *father*. Este campo almacena el identificador del comentario al cual se responde, en caso de tratarse de una respuesta; por lo tanto, no todos los comentarios contienen un valor en dicho atributo.

Tabla 9: Cabecera del nuevo archivo CSV para realizar el análisis de sentimientos.
 Fuente: Elaboración propia.

post_id	pk	user_id	full_name	did_report_as_spam	created_at_utc	is_created_by_media_owner	text	father
---------	----	---------	-----------	--------------------	----------------	---------------------------	------	--------

En cuanto a las *captions* de las publicaciones, se aplicó un proceso de preprocesamiento que consistió en las siguientes etapas:

1. Eliminación de tildes, símbolos diacríticos y emojis.
2. Conversión de todo el texto a minúsculas.

3. Eliminación de stopwords.
4. Eliminación de registros duplicados.

En el caso de los comentarios, no se aplicó un preprocesamiento exhaustivo, ya que la herramienta utilizada, *pysentimiento*, al estar optimizada para datos provenientes de redes sociales, incorpora internamente los procesos de normalización y limpieza textual. Esto se puede confirmar al revisar su código fuente, en especial el archivo `preprocessing.py`¹⁰. La única transformación aplicada a los comentarios fue la eliminación de duplicados.

3.5. Modelado: primera iteración

La selección de modelos para la etapa de modelado en un proceso de minería de datos no solo responde a criterios de disponibilidad o facilidad de uso, sino que debe estar orientada por la naturaleza del problema, las características de los datos y los objetivos del análisis. En esta memoria, centrada en analizar los sentimientos expresados en comentarios de publicaciones sobre salud mental en una red social, se optó por la utilización de dos modelos preentrenados de análisis de sentimientos: *pysentimiento* y la API de *Google Natural Language*.

El modelo principal seleccionado fue *pysentimiento* [Pérez *et al.*, 2023], una biblioteca desarrollada específicamente para tareas de procesamiento de lenguaje natural (PLN) aplicadas a textos en redes sociales, con un fuerte enfoque en el idioma español. Esta herramienta se basa en modelos de aprendizaje profundo entrenados sobre *corpus* sociales reales, lo cual le otorga una ventaja comparativa significativa al momento de analizar textos breves, informales y cargados de expresividad subjetiva, como es habitual en los comentarios de usuarios en plataformas digitales. Además, al estar desarrollada en Python y disponible como proyecto *open-source*, su integración en el flujo de procesamiento fue directa, transparente y sin costos asociados. Si bien *pysentimiento* es una biblioteca que emplea distintos modelos, los cuales selecciona según parámetros definidos, en esta memoria se utilizó *pysentimiento* directamente, y no un modelo específico de manera aislada. Por lo tanto, al hacer referencia al “modelo”, se refiere al modelo subyacente que *pysentimiento* utiliza para la tarea correspondiente.

Una de las fortalezas más destacadas de *pysentimiento* radica en su optimización para el análisis de texto en español, evitando así la necesidad de traducción de los datos —un proceso que podría introducir ambigüedades o pérdida de matices en los textos originales—. A diferencia de otros modelos ampliamente conocidos, como VADER (*Valence Aware Dictionary and sEntiment Reasoner*) o TextBlob, que están diseñados principalmente para el inglés o que utilizan reglas lingüísticas generales [Di Bella, 2025], *pysentimiento* incorpora los pa-

¹⁰<https://github.com/pysentimiento/pysentimiento/tree/master/pysentimiento>

trones del habla propios de las redes sociales en español, que pueden incluir expresiones coloquiales, abreviaciones, emojis o neologismos comunes.

Una decisión relevante en esta etapa fue no utilizar otros modelos populares como VADER o TextBlob, a pesar de su disponibilidad y uso extendido en proyectos de análisis de sentimientos. Aunque ambas herramientas forman parte del conjunto de soluciones de PLN, presentan limitaciones importantes en el contexto específico de esta memoria. En primer lugar, tanto VADER como TextBlob fueron desarrolladas principalmente para el idioma inglés, y sus versiones adaptadas al español implican traducir previamente el texto. Esto limita significativamente su precisión en tareas que dependen de matices lingüísticos y culturales, como el análisis de emociones y opiniones en español.

Además, estos modelos están basados en enfoques léxicos y reglas estáticas, lo que implica una menor capacidad de adaptación frente a variaciones del lenguaje informal presentes en redes sociales, como abreviaciones, emoticones, sarcasmo o neologismos. Por ejemplo, VADER, aunque fue diseñado para contenido de redes sociales, utiliza un diccionario predefinido con reglas gramaticales fijas, lo cual restringe su capacidad de clasificar el texto contextual, a pesar de que en su diccionario se incluyen los emojis. TextBlob, por su parte, ofrece una gama más amplia de funcionalidades de PLN, pero su modelo de análisis de sentimientos depende también de reglas simples y está optimizado para textos largos y formales no estando diseñado para analizar textos en español.

En contraste, *pysentimiento* incorpora modelos de aprendizaje profundo basados en transformers (como BERT y RoBERTa), capaces de captar el contexto semántico y emocional de los mensajes de forma mucho más precisa. Al haber sido entrenado con un enfoque en la minería de datos para redes sociales, refleja con mayor fidelidad el tipo de lenguaje analizado en esta memoria.

Complementariamente, se utilizó la API de *Google Natural Language*, una herramienta de carácter comercial que proporciona acceso a modelos de PLN entrenados con grandes volúmenes de datos heterogéneos. Esta API representa una solución robusta y ampliamente validada en contextos empresariales, y permite realizar análisis semánticos complejos a través de solicitudes remotas. Entre sus capacidades destacan el análisis de sentimientos, el reconocimiento de entidades, la clasificación de contenido y el análisis sintáctico, lo que la convierte en una herramienta versátil para contrastar los resultados obtenidos por medios alternativos. Su utilización en esta memoria permitió enriquecer la interpretación de los comentarios desde una perspectiva metodológica distinta, y sirvió como punto de comparación con el modelo principal.

La combinación de ambos modelos no solo obedece a un criterio de validación cruzada, sino que responde también a una estrategia de complementariedad metodológica. Mientras que *pysentimiento* representa un enfoque especializado, contextualizado y enfocado en el dominio social hispanohablante, la API de Google aporta una visión más generalista y consolidada en entornos productivos. Esta dualidad permite observar coincidencias y divergencias en la clasificación de los sentimientos, lo cual contribuye a identificar posibles sesgos, evaluar la

robustez del análisis y mejorar la confiabilidad general de los resultados y sus interpretaciones.

En suma, la elección de estos dos modelos se fundamenta en su capacidad conjunta para abordar los desafíos propios del lenguaje en redes sociales, su adaptabilidad técnica al entorno de trabajo de esta memoria, y su valor complementario para garantizar una evaluación más integral y precisa de las percepciones expresadas en los comentarios analizados.

3.5.1. Resultados

Para el análisis de sentimientos usando *pysentimiento* en Python se empleó la biblioteca *pysentimiento*. La inicialización del modelo se realizó mediante la función *create_analyzer*, a la cual se le asignaron los parámetros *task='sentiment'* y *lang='es'*, indicando la tarea de clasificación de sentimientos en idioma español. Esta función retorna un objeto de la clase *AnalyzerForSequenceClassification*, que encapsula un modelo de clasificación basado en transformers preentrenados, seleccionado según los parámetros indicados.

El método *predict()* de esta clase fue utilizado para analizar cada texto individual (en este caso, los comentarios). Este método recibe un string como entrada y retorna un objeto de tipo *AnalyzerOutput*, que contiene:

- **output**: el tipo de sentimiento predicho (POS, NEG o NEU).
- **probas**: un diccionario con la probabilidad asignada a cada sentimiento por el modelo.

El formato de salida sigue la siguiente estructura: *output=POS, probas={POS: 0.998, NEG: 0.002, NEU: 0.000}*.

La Tabla 10 describe el significado de cada etiqueta de sentimiento:

Tabla 10: Etiquetas de sentimiento y su significado según el modelo *pysentimiento*.
 Fuente: Elaboración propia.

Etiqueta	Palabra	Significado
POS	Positive	El comentario expresa una opinión o sentimiento positivo.
NEG	Negative	El comentario refleja una opinión o sentimiento negativo.
NEU	Neutral	El comentario no expresa una carga emocional clara; es neutral.

En la Tabla 11 se presenta la estructura del archivo CSV en el cual se guardaron los resultados del análisis de sentimientos realizado con *pysentimiento*. Los campos que son nuevos con

respecto a los archivos mencionados previamente, son *sentiment* y *polaridad*. A continuación se detalla el significado de estos campos.

- **post_id**: Corresponde al identificador del *post* al cual pertenece el comentario.
- **comment_id**: Corresponde al identificador del comentario.
- **father**: Corresponde al identificador del comentario padre del comentario actual.
- **comment**: Corresponde al contenido del comentario.
- **sentiment**: Representa la clase de sentimiento predicha por el modelo, la cual puede tomar uno de los siguientes valores: *positivo*, *neutro* o *negativo*, en correspondencia con las etiquetas entregadas por el modelo (POS, NEU, NEG).
- **polaridad**: Corresponde a la probabilidad asignada por el modelo a la clase de sentimiento seleccionada. Esta se expresa como un valor decimal en el rango [0, 1].

Tabla 11: Cabecera del archivo CSV que guarda los resultados del análisis de sentimientos hecho con *pysentimiento*.

Fuente: Elaboración propia.

post_id	comment_id	father	comment	sentiment	polaridad
---------	------------	--------	---------	-----------	-----------

Para el análisis de sentimientos utilizando la API de *Google Natural Language*, se empleó la biblioteca *google.cloud* en Python, importando específicamente el módulo *language_v1*. A partir de este módulo, se instanció el cliente del servicio mediante el método `LanguageServiceClient()`.

Luego, se construyó un objeto del tipo `Document` a través del mismo módulo, definiendo los siguientes parámetros: `content=text`, `type=language_v1.Document.Type.PLAIN_TEXT` y `language="es"`. En este contexto, `text` representa el contenido del comentario a analizar, `type` especifica el tipo de texto (en este caso, texto plano), y `language` indica el idioma del texto (español). Esta configuración permite generar una estructura de documento (objeto `Document`) válida para el modelo.

A continuación, se aplicó el método `analyze_sentiment()` sobre el cliente, pasando como parámetro el documento previamente construido a través del diccionario `request="document": document`. El resultado devuelto contiene múltiples componentes, entre ellos `document_sentiment`, que corresponde al análisis general del sentimiento del texto completo, proporcionando una puntuación de polaridad y un valor de magnitud. El sentimiento tiene la siguiente estructura:

- **score:** es un valor que oscila entre -1.0 y +1.0. Indica la polaridad general del sentimiento dentro del texto o la oración. Valores positivos (cercaos a +1.0): sugieren un sentimiento positivo; cuanto más cerca de +1.0, más fuerte es el sentimiento positivo. Valores negativos (cercaos a -1.0): indican un sentimiento negativo; cuanto más cerca de -1.0, más fuerte es el sentimiento negativo. Valores cercanos a 0.0 pueden indicar un sentimiento neutral, pero también podrían significar que el texto contiene sentimientos tanto positivos como negativos que se neutralizan entre sí. Para distinguir estos casos, se debe observar la magnitud.
- **magnitudo:** es un valor no negativo que va de 0.0 a infinito. Representa la fuerza o intensidad general de la emoción expresada en el texto o la oración, independientemente de su polaridad (positiva o negativa). Valores bajos (cercaos a 0.0) indican que hay poca intensidad emocional en el texto; un texto puramente objetivo o neutral podría tener una magnitud baja. Valores altos: sugieren que se expresa una cantidad significativa de emoción en el texto; Un texto con muchos adjetivos calificativos o expresiones emocionales tendrá una magnitud alta. Un texto con un *score* cercano a 0.0 y una *magnitudo* alta sugiere que hay sentimientos mixtos presentes en el texto, donde tanto las opiniones positivas como negativas son fuertes. En cambio, un *score* cercano a 0.0 y una *magnitudo* baja indica un texto genuinamente neutral o con muy poca carga emocional.

En la Tabla 12 se presenta la estructura del archivo CSV en el cual se guardaron los resultados del análisis de sentimientos realizado con la API de *Google Natural Language*. Los campos que son nuevos con respecto a los archivos mencionados previamente, son *score* y *magnitudo*, previamente ya explicados.

Tabla 12: Cabecera del archivo CSV que guarda los resultados del análisis de sentimientos hecho con la API de *Google Natural Language*.

Fuente: Elaboración propia.

post_id	comment_id	father	comment	score	magnitudo
---------	------------	--------	---------	-------	-----------

Para el análisis de emociones realizado con *pysentimiento*, se empleó el mismo procedimiento utilizado en el análisis de sentimientos, con la única diferencia de que el parámetro *task* se establece con el valor *emotion*. El resto de la configuración, incluyendo el idioma y la utilización del método *predict()*, se mantiene sin modificaciones.

El resultado de la predicción entrega una etiqueta correspondiente a la emoción predominante en el texto, junto con un diccionario de probabilidades asociadas a cada clase emocional.

- **output:** el tipo de emoción predicha (*others* , *joy*, *sadness*, *anger*, *surprise*, *disgust*, *fear*).

- **probas**: un diccionario con la probabilidad asignada a cada emoción por el modelo.

El formato de salida sigue la siguiente estructura: `output=joy, probas={'others': 0.1393018215894699, 'joy': 0.8356634974479675, 'sadness': 0.006904544308781624, 'anger': 0.0015291842864826322, 'surprise': 0.012434794567525387, 'disgust': 0.0022084543015807867, 'fear': 0.001957688946276903}`.

Los resultados de este análisis se guardaron en un archivo CSV con la estructura ilustrada en la Tabla 13, donde `emoción` toma el valor de la emoción con mayor probabilidad entregada por el modelo (`output`) y `probabilidades` guarda el diccionario con las probabilidades para cada emoción.

Tabla 13: Cabecera del archivo CSV que guarda los resultados del análisis de emociones hecho con `pysentimiento`.

Fuente: Elaboración propia.

post_id	comment_id	father	comment	emoción	probabilidades
---------	------------	--------	---------	---------	----------------

3.6. Evaluación: primera iteración

La etapa de Evaluación es un paso clave dentro de CRISP-DM, cuyo propósito principal es examinar los resultados obtenidos durante el modelado desde una perspectiva integral que considere tanto aspectos técnicos como prácticos y estratégicos. En esta sección se realizará una revisión crítica de los dos modelos preentrenados para analizar sentimientos en comentarios extraídos de publicaciones sobre salud mental en Instagram, destacando no solo el desempeño cuantitativo, sino también el valor desde una perspectiva académica y social más amplia.

Debido a la naturaleza del conjunto de datos utilizado, el cual no posee etiquetas preexistentes, se adopta una evaluación mixta basada en la combinación de un análisis cuantitativo mediante la comparación manual de una muestra representativa de comentarios y una revisión cualitativa exploratoria para determinar la coherencia y utilidad práctica de los resultados obtenidos. De esta manera, se pretende identificar fortalezas, limitaciones y oportunidades de mejora de los modelos aplicados.

En primera instancia, es necesario establecer el umbral de clasificación que permita interpretar los resultados obtenidos a partir del análisis de sentimientos mediante la API de *Google Natural Language*. Como se expuso en la sección anterior, dicha herramienta no asigna categorías de sentimiento de manera predeterminada, sino que entrega un puntaje continuo (*score*) que refleja la polaridad del texto. En este contexto, corresponde al investigador determinar los valores límite que delimiten las clases de análisis —positivo, negativo o neutro—,

con el objetivo de asegurar una interpretación consistente y justificada según criterios metodológicos definidos.

Para definir este umbral de clasificación, se evaluaron distintos valores para el parámetro *score*. Aunque valores más altos (como ± 0.5) ofrecían mayor precisión en las coincidencias exactas con el modelo *pysentimiento*, como se puede observar en la Tabla 14, tendían a generar una sobrerrepresentación de sentimientos neutros. En cambio, el umbral de ± 0.25 permitió capturar una mayor proporción de comentarios con sentimientos definidos, sin una pérdida significativa de concordancia general, como se puede observar en la Tabla 15. Por esta razón, se optó por utilizar *score* mayor a 0.25 para etiquetar comentarios como positivos, *score* menor a -0.25 como negativos y aquellos dentro del rango $[-0.25, 0.25]$ como neutros. Asimismo, la documentación oficial de la API señala que un *score* de 0.2 corresponde a una emoción levemente positiva, por lo que establecer el umbral en 0.25 resulta coherente con los lineamientos propuestos como ejemplo por la herramienta.

Tabla 14: Comparación entre las clasificaciones de sentimientos de Google y *pysentimiento* (umbral 0.5)

Fuente: Elaboración propia.

	negativo (Google)	neutro (Google)	positivo (Google)
negativo (<i>pysentimiento</i>)	256	2,478	140
neutro (<i>pysentimiento</i>)	41	5,288	1,010
positivo (<i>pysentimiento</i>)	0	1,151	3,707

Tabla 15: Comparación entre las clasificaciones de sentimientos de Google y *pysentimiento* (umbral 0.25)

Fuente: Elaboración propia.

	negativo (Google)	neutro (Google)	positivo (Google)
negativo (<i>pysentimiento</i>)	628	1,820	426
neutro (<i>pysentimiento</i>)	120	3,397	2,822
positivo (<i>pysentimiento</i>)	7	430	4,421

3.6.1. Métricas de rendimiento

Etiquetado manual

Para la evaluación del desempeño de los modelos de análisis de sentimientos, se utilizó una muestra de 357 comentarios provenientes de una base previamente muestreada de manera balanceada según las predicciones del modelo *pysentimiento*. Cada comentario fue etiquetado manualmente como positivo, negativo o neutro, generando así un conjunto de referencia confiable. Posteriormente, con el objetivo de eliminar posibles sesgos introducidos por la distribución de predicciones iniciales, se construyó una nueva muestra balanceada según las

etiquetas reales asignadas manualmente. Esta muestra final quedó conformada por 119 comentarios por clase real, lo que permitió realizar una evaluación equitativa y comparativa del rendimiento de los modelos *pysentimiento* y *Google Natural Language*, bajo condiciones de representación uniforme entre clases.

Análisis e interpretación de resultados

A continuación, para obtener las matrices de confusión y métricas de rendimiento, se utilizaron las funciones `confusion_matrix` y `classification_report` de la biblioteca `sklearn.metrics`.

Tabla 16: Matriz de confusión - *pysentimiento*
 Fuente: Elaboración propia.

	positivo	negativo	neutro
positivo	88	3	28
negativo	0	116	3
neutro	1	27	91

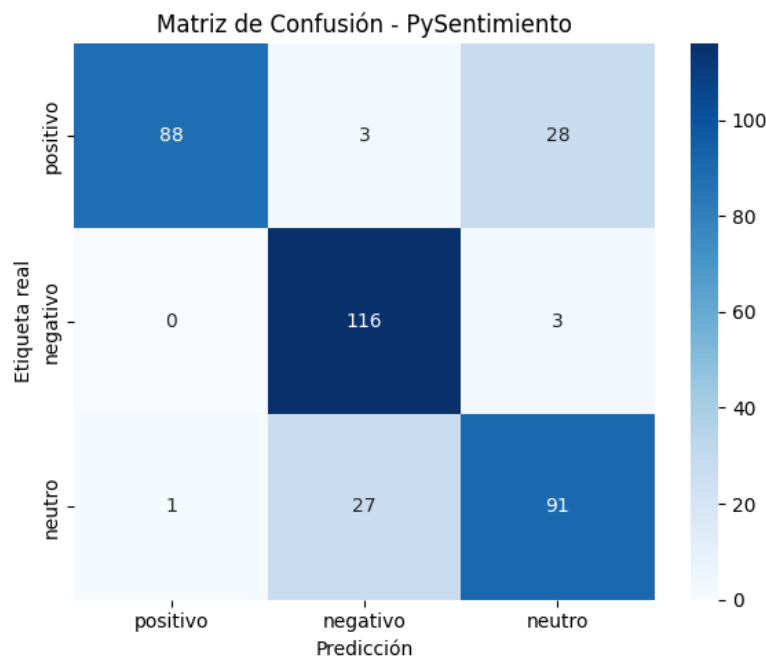


Figura 9: Matriz de confusión correspondiente al modelo de análisis de sentimientos *pysentimiento* evaluado sobre una muestra balanceada por clase real.

Fuente: Elaboración propia.

La Figura 9 muestra la matriz de confusión para los resultados del análisis de sentimientos con *pysentimiento*. En el eje vertical se presentan las etiquetas reales, mientras que en el eje horizontal se muestran las predicciones del modelo. La diagonal principal indica las clasificaciones correctas.

Tabla 17: Matriz de confusión - *Google Natural Language*
 Fuente: Elaboración propia.

	positivo	negativo	neutro
positivo	94	0	25
negativo	15	32	72
neutro	40	11	68

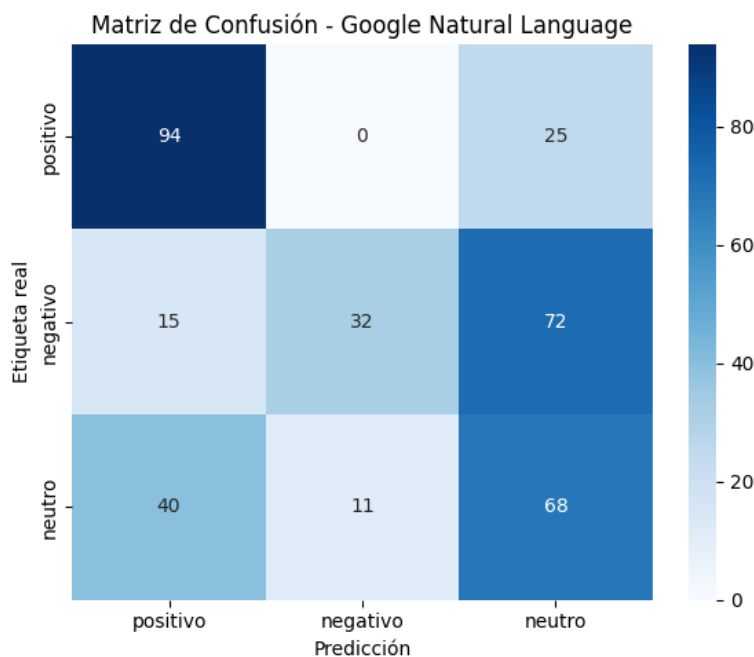


Figura 10: Matriz de confusión correspondiente al modelo de análisis de sentimientos de la API de *Google Natural Language* evaluado sobre una muestra balanceada por clase real.

Fuente: Elaboración propia.

La Figura 10 muestra la matriz de confusión para los resultados del análisis de sentimientos con *pysentimiento*. En el eje vertical se presentan las etiquetas reales, mientras que en el eje horizontal se muestran las predicciones del modelo. La diagonal principal indica las clasificaciones correctas.

Tabla 18: Métricas de evaluación - pysentimiento
 Fuente: Elaboración propia.

Clase	Precisión	Recall	F1-score	Soporte
positivo	0.989	0.739	0.846	119
negativo	0.795	0.975	0.875	119
neutro	0.746	0.765	0.755	119
Promedio macro	0.843	0.826	0.826	357
Promedio ponderado	0.843	0.826	0.826	357

Tabla 19: Métricas de evaluación - Google Natural Language
 Fuente: Elaboración propia.

Clase	Precisión	Recall	F1-score	Soporte
positivo	0.631	0.790	0.701	119
negativo	0.744	0.269	0.395	119
neutro	0.412	0.571	0.479	119
Promedio macro	0.596	0.543	0.525	357
Promedio ponderado	0.596	0.543	0.525	357

Al comparar el desempeño de los modelos *pysentimiento* y *Google Natural Language* sobre la muestra balanceada por clase real (119 ejemplos por categoría), se observan diferencias sustanciales en cuanto a precisión global y desempeño por clase.

El modelo *pysentimiento* alcanzó una exactitud (*accuracy*) global del 82,6%, destacándose por un elevado *recall* para la clase “negativo” (0.975), lo que indica una alta capacidad para detectar correctamente este tipo de comentarios. Para las clases “positivo” y “neutro”, si bien el *recall* fue más moderado (0.739 y 0.765 respectivamente), la precisión se mantuvo alta, especialmente en la clase “positivo” (0.989). Estas cifras se reflejan en un promedio macro (macro avg) de F1-score de 0.826, lo que muestra un buen equilibrio entre las clases, considerando que la muestra estaba perfectamente balanceada.

Por otro lado, el modelo de *Google Natural Language* obtuvo una exactitud significativamente menor (54,3%), con un desempeño particularmente bajo en la clase “negativo”, donde el *recall* fue de apenas 0.269. Si bien logró una mejor cobertura en la clase “positivo” (*recall* de 0.790), el rendimiento en “neutro” también fue limitado (F1-score de 0.479). El F1-score macro promedio fue de 0.525, lo que evidencia una menor consistencia en su capacidad de clasificación entre clases.

Al observar las matrices de confusión correspondientes a ambos modelos (ver Figuras 9 y 10), se evidencian diferencias significativas en su capacidad para clasificar correctamente los comentarios según el sentimiento. En el caso de *pysentimiento*, se observa una diagonal principal claramente definida, lo cual indica que la mayoría de las predicciones coinciden con las etiquetas reales. En particular, destaca la celda correspondiente a la clase negativo, donde

se registran 116 aciertos sobre 119 casos reales, lo que confirma su elevada sensibilidad (*recall*) en esta categoría. Aunque existe cierta confusión entre positivo y neutro, esta es moderada y no compromete gravemente la precisión general del modelo. En resumen, la visualización refleja un modelo bien calibrado, con errores localizados y razonables.

Por el contrario, la matriz de confusión del modelo de Google presenta una distribución mucho más dispersa, con numerosos errores fuera de la diagonal. Aunque la clase positivo muestra una buena tasa de aciertos (94 de 119), en las clases negativo y neutro el modelo incurre en una gran cantidad de errores. Específicamente, los casos negativos tienden a ser confundidos con neutros o incluso con positivos, lo que revela una mayor ambigüedad en los límites entre clases. La dispersión de las celdas en la matriz indica que el modelo no logra distinguir de forma clara los matices del lenguaje emocional, especialmente en textos con contenido más crítico o ambiguo.

Visualmente, las diferencias entre ambas matrices no solo confirman los resultados cuantitativos reportados en las métricas (precisión, *recall* y F1-score), sino que permiten identificar patrones de error. Mientras que *pysentimiento* concentra sus errores entre clases similares, el modelo de Google presenta errores más desbalanceados, lo cual compromete su utilidad en contextos donde la detección precisa de sentimientos negativos o neutros es especialmente relevante y si bien en este trabajo no es crítico este error, es muy importante tenerlo en cuenta.

En términos generales, *pysentimiento* demostró un rendimiento superior tanto a nivel global como por clase, especialmente en la identificación de comentarios negativos, mientras que el modelo de Google mostró mayores dificultades para distinguir correctamente las tres categorías, con una fuerte tendencia a confundir comentarios negativos con neutros o positivos.

¿Por qué podrían haberse producido diferencias tan marcadas en el rendimiento? El resultado puede parecer sorprendente, considerando que se está comparando un modelo de código abierto con uno desarrollado por una empresa de alcance global y con vastos recursos tecnológicos, como Google. Una posible explicación radica en los *corpus* de entrenamiento utilizados. En el caso de *pysentimiento*, se sabe que ha sido optimizado específicamente para el análisis de textos provenientes de redes sociales, lo que le otorga una ventaja en este contexto particular. En contraste, el modelo de *Google Natural Language*, si bien declara ser aplicable a múltiples dominios, no parece estar especialmente ajustado a los patrones lingüísticos propios de las redes sociales. Esta falta de especialización podría explicar su menor desempeño frente a un modelo entrenado con foco explícito en ese tipo de contenido.

Por otro lado, es posible que el modelo de *Google Natural Language* tenga un buen desempeño en el análisis de sentimientos de datos provenientes de redes sociales, pero enfocado en tipos de usuarios o contextos distintos a los abordados en esta memoria. Dado que Google no publica con detalle el *corpus* de entrenamiento utilizado, no se puede descartar que el modelo haya sido optimizado para dominios más generales o comerciales, como reseñas de productos, atención al cliente o interacciones corporativas. Esta diferencia en el enfoque

podría explicar su menor rendimiento frente a *pysentimiento*, especialmente en textos con temáticas sensibles o cargadas emocionalmente, como los relacionados con salud mental. Este fenómeno es consistente con lo que en la literatura puede encontrarse, en inglés, como *domain shift problem* o *cross-domain generalization*. Estos términos hacen referencia a la dificultad que enfrentan los modelos al ser aplicados en contextos diferentes a aquellos para los que fueron entrenados. En particular, un modelo entrenado sobre ciertos tipos de lenguaje, temáticas o comunidades puede experimentar una disminución en su rendimiento al enfrentarse a datos provenientes de entornos con características discursivas distintas.

Dado que los resultados obtenidos por el modelo de *Google Natural Language* no fueron óptimos, y considerando que *pysentimiento* demostró un desempeño superior tanto en términos comparativos como en las métricas de evaluación, se decidió continuar el análisis exclusivamente en base a los resultados de este último modelo.

Tabla 20: Evolución mensual de los sentimientos en los comentarios, clasificados mediante *pysentimiento*.

Fuente: Elaboración propia.

Mes	positivo	neutro	negativo
2022-04	9	11	3
2022-05	1	1	0
2022-06	0	1	0
2022-10	19	32	11
2022-11	7	10	1
2022-12	250	271	206
2023-01	267	357	267
2023-02	269	316	129
2023-03	239	336	180
2023-04	260	319	303
2023-05	145	258	151
2023-06	152	182	107
2023-07	80	106	26
2023-08	257	318	133
2023-09	253	302	112
2023-10	525	495	192
2023-11	264	327	102
2023-12	294	230	78
2024-01	224	317	207
2024-02	276	676	161
2024-03	372	777	126
2024-04	425	431	225
2024-05	270	266	154

3.6.2. Resultados: Análisis de Sentimientos

En las Tablas 20 y 21 se muestran resúmenes de los resultados obtenidos del análisis de sentimientos realizado con *pysentimiento*. A partir de los datos presentados, se observa un incremento sostenido en la cantidad de comentarios clasificados por sentimiento desde diciembre de 2022, alcanzando un primer punto alto en octubre de 2023, con 525 comentarios positivos, 495 neutros y 192 negativos. A lo largo del periodo analizado, predominan los comentarios de carácter positivo y neutro, mientras que los negativos se mantienen en menor proporción relativa.

Tabla 21: Distribución porcentual mensual de los sentimientos en los comentarios, clasificados mediante *pysentimiento*.

Fuente: Elaboración propia.

Mes	positivo (%)	neutro (%)	negativo (%)
2022-04	39	48	13
2022-05	50	50	0
2022-06	0	100	0
2022-10	31	52	18
2022-11	39	56	6
2022-12	34	37	28
2023-01	30	40	30
2023-02	38	44	18
2023-03	32	45	24
2023-04	29	36	34
2023-05	26	47	27
2023-06	34	41	24
2023-07	38	50	12
2023-08	36	45	19
2023-09	38	45	17
2023-10	43	41	16
2023-11	38	47	15
2023-12	49	38	13
2024-01	30	42	28
2024-02	25	61	14
2024-03	29	61	10
2024-04	39	40	21
2024-05	39	39	22

3.6.3. Resultados: Análisis de Emociones

Los resultados del análisis de emociones realizado con *pysentimiento* se muestran en la Tabla 22, la cual presenta la distribución de emociones detectadas en los comentarios analizados. Se observa que la categoría *others* concentra la mayor cantidad de casos con 8.427 registros, lo cual indica que una proporción significativa de los comentarios no fue asociada de manera predominante a ninguna de las emociones específicas consideradas por el modelo.

Tabla 22: Distribución de emociones detectadas
Fuente: Elaboración propia.

Emoción	Cantidad
others	8.427
joy	4.255
anger	1.000
sadness	843
surprise	134
fear	39
disgust	18

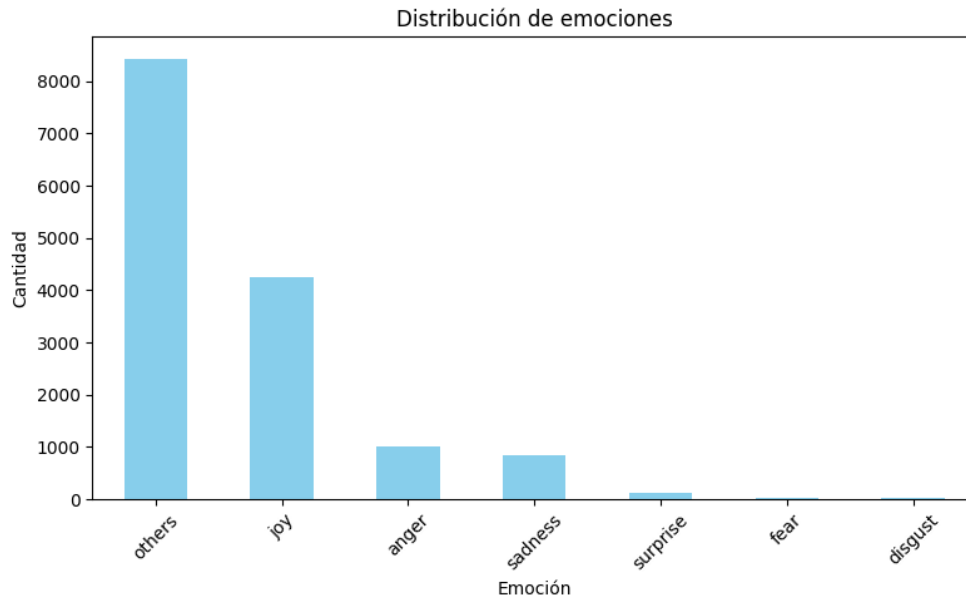


Figura 11: Frecuencia de las emociones clasificadas por *pysentimiento*.
Fuente: Elaboración propia.

Dado que una proporción considerable de los resultados fue clasificada como *others*, lo que sugiere una alta incertidumbre o ambigüedad en dichas predicciones, se decide reiterar la etapa de Modelado con el objetivo de mejorar la calidad y precisión de las clasificaciones obtenidas.




3.7. Modelado: segunda iteración

Durante la fase de evaluación se observaron limitaciones en el desempeño del modelo *py-sentimiento* en la clasificación por emociones, particularmente en su tendencia a clasificar como *others* un número significativo de comentarios que, desde una perspectiva humana, manifestaban una carga emocional evidente a través del uso de emojis. Esta situación motivó una revisión crítica de los resultados y, en consecuencia, una iteración metodológica.

Siguiendo la flexibilidad que propone el enfoque CRISP-DM, se aprovechó la posibilidad de iterar desde la fase de evaluación hacia la fase de modelado, con el objetivo de mejorar la calidad de las predicciones. Esta segunda iteración no implicó reentrenar el modelo base, sino integrar una heurística complementaria basada en el análisis semántico de emojis.

3.7.1. Corrección heurística de emociones a partir de emojis

Durante el proceso de clasificación emocional de los comentarios mediante el modelo *py-sentimiento* (`task="emotion"`), se detectaron múltiples casos en los que el modelo devolvía la categoría *others*. Esta clase es utilizada por el modelo para representar comentarios cuya carga emocional no se asocia directamente con ninguna de las emociones básicas reconocidas (*joy*, *sadness*, *anger*, *fear*, *disgust*, *surprise*). Sin embargo, al revisar manualmente varios de estos casos, se observó que muchos incluían emojis con una carga emocional evidente desde una perspectiva humana.

Con el fin de mejorar la cobertura del análisis emocional y reducir falsos negativos en los resultados, se implementó una estrategia de corrección heurística basada en emojis. Este enfoque consistió en construir un diccionario manual que asocia ciertos emojis con emociones básicas. Por ejemplo, el emoji  fue asociado a *sadness*, mientras que , y  fueron asignados a *joy*. Este diccionario se generó a partir de un análisis exploratorio de frecuencia de uso de emojis en los comentarios y se almacenó en un archivo independiente (`clasificacion_emojis.csv`) con la estructura `emoji,count,emotion`. Para emojis cuya frecuencia era menor a 5 apariciones, no se les consideró en el diccionario final.

Dado que la lista de emojis contenía un total de 159 casos y que el etiquetado manual puede estar sujeto a sesgos subjetivos, se optó por utilizar tres modelos distintos como apoyo en la asignación de emociones a cada emoji. A todos ellos se les entregó el *prompt* mostrado a continuación.

Objetivo:

Clasifica cada emoji del dataset en una de las siguientes emociones básicas: *joy* (alegría), *surprise* (sorpresa), *sadness* (tristeza), *anger* (enojo), *disgust* (disgusto) o *fear* (miedo).

Instrucciones:

Asigna una sola emoción principal por emoji (elige la más representativa).

Considera el uso común de cada emoji en redes sociales y mensajería.

Emojis ambiguos: Si un emoji puede tener múltiples significados, elige el más frecuente (ejemplo: 😊 puede ser “*joy*” por nerviosismo positivo o “*surprise*” por incomodidad; en este caso, “*joy*” es más común).

Formato de salida esperado:

emoji,count,emotion

❤️ ,3319,*joy*

👏 ,2624,*joy*

Ejemplos de referencia:

❤️ → *joy* (amor, positividad)

😭 → *sadness* (llanto, tristeza)

Genera el dataset completo con la columna *emotion* añadida, siguiendo las pautas anteriores.

Los modelos utilizados fueron:

- GPT-4o
- DeekSeep-V3
- Gemini 2.5 Pro (Preview)

Finalmente, se compararon los resultados obtenidos por los tres modelos. Tras un análisis cualitativo, se determinó que las clasificaciones generadas por DeekSeep-V3 eran las más consistentes y adecuadas. Por esta razón, se optó por utilizar dichas etiquetas como resultado final del proceso de asignación emocional por emoji. Las primeras 15 asignaciones se presentan en la Tabla 23.

Tabla 23: Frecuencia y emoción asignada a cada emoji para los 15 más frecuentes.
 Fuente: Elaboración propia.

Emoji	Frecuencia	Emoción
❤️	3319	joy
👏	2624	joy
🙌	1632	joy
😍	757	joy
🔥	578	joy
😞	554	joy
😞	552	sadness
💜	322	joy
🙏	293	joy
✨	281	joy
😡	221	anger
😊	214	joy
😘	163	joy
😏	157	joy

Posteriormente, se aplicó una función sobre cada comentario en el que *pysentimiento* había predicho *others*. Esta función analizaba los emojis presentes en el texto y, si encontraba alguno registrado en el diccionario, asignaba su emoción correspondiente. En caso contrario, el valor *others* se mantenía. La emoción ajustada se almacenó en una nueva columna llamada *heurística*, la cual coincide con la emoción original (*emocion*) en todos los casos donde *pysentimiento* predijo una emoción válida, y difiere solamente en aquellos casos donde *others* fue corregido por un emoji emocional reconocible.

Ejemplo aplicado: Para el comentario “*info necesaria* 🍷🍷”, el modelo *pysentimiento* asignó la emoción *others* al no detectar un patrón lingüístico explícitamente emocional. No obstante, gracias a la heurística implementada, se identificó el emoji 🍷, asociado a *joy*, y se asignó dicha emoción como corrección en la columna *heurística*.

Esta estrategia permitió extender el alcance del análisis emocional a casos breves o implícitos donde el modelo no lograba una clasificación precisa, especialmente en comentarios dominados por contenido visual (emojis) más que verbal.

Este proceso no sustituye el modelo original, sino que lo complementa como parte del refinamiento lógico del sistema de clasificación. Así, se demuestra cómo la fase de evaluación puede revelar oportunidades de mejora que justifican una iteración hacia fases anteriores del ciclo CRISP-DM, consolidando un enfoque flexible y adaptativo.

3.8. Evaluación: segunda iteración

Luego de realizar una segunda iteración en la etapa de Modelado, se obtuvieron mejoras significativas en la clasificación de emociones. En particular, se observa una reducción considerable en la cantidad de casos etiquetados como *others*, y un aumento en las emociones clasificadas positivamente, como *joy*, lo cual sugiere una mejor adaptación del modelo a las características del conjunto de datos. La distribución resultante se muestra en la Tabla 24.

Tabla 24: Distribución de emociones detectadas tras segunda iteración
 Fuente: Elaboración propia.

Emoción	Cantidad
<i>joy</i>	7891
<i>others</i>	4375
<i>anger</i>	1047
<i>sadness</i>	973
<i>surprise</i>	300
<i>disgust</i>	64
<i>fear</i>	58

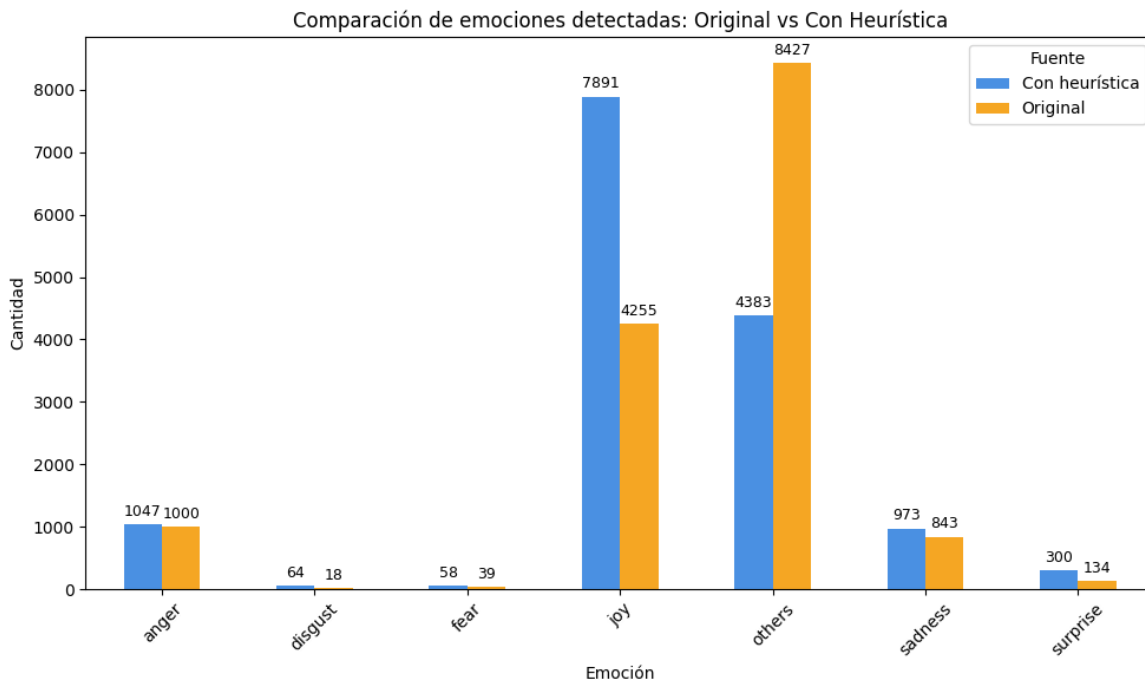


Figura 12: Comparación de la distribución de emociones detectadas por el modelo original y el ajustado con heurísticas.

Fuente: Elaboración propia.

Tabla 25: Ejemplos de comentarios cuya clasificación de emoción fue modificada por la heurística.

Fuente: Elaboración propia.

Comentario	Emoción
amoo ❤️ para mi ideal un paseo viendo los lindos colores de otoño y tomando un cafecito 😞	joy
😍😍	joy
@user 🤍🤍🤍🤍	joy
ame ❤️	joy
@user 💜💚👊	joy
siempre recuerdo de cuando siendo niño, si hacia algo mal me subian al auto y me dejaba tirado en la playa... jajaja cómo que no es normal eso? 🤔	surprise
😏	surprise
no entiendo la mezcla de temas 🤔	surprise
@user no pensé que tenía que explicar hasta el ejemplo. y me da paja hacerlo en realidad. cada quien entiende lo que quiere entender. o lo que puede 🙄	surprise
el deporte es una actividad consentida, las partes que participan lo hacen por voluntad....! se imaginan un futbolista o un tenista obligado en la cancha...???? 🤔🤔🤔🤔	surprise
@user 🙄💜	sadness
punpun 💔	sadness
@user todos necesitamos un tío iroh en nuestras vidas 🙄 jajaja	sadness
@user a mi tambien, pero tengo otro compromiso ese dia 🙄	sadness
@user 🙄🤔	sadness
muy buen <i>post</i> hoy me paso 🙄	sadness
ya no sé cuántas veces me han dicho facha 🙄	anger
@user 🙄💚💜	anger
agenda 2030 y sus horrores 🤔	anger
siempre encuentran un hueco legal 🙄	anger
quien como ellos, yo supongo y asumo cosas todo el dia 🙄	anger
@user 🙄🙄🙄	disgust
😞😞	disgust
esas uniones vienen de hace rato 🤔	disgust
mucho tiempo creí que era tonta porque al ponerme nerviosa me equivocaba mucho y me bloqueaba por completo, pero era esto 🤔	fear
@user kernberg explica la leve diferencia que tienen, lo cierto es que son muy parecidos 🙄	fear
¿qué hacer cuando el gatillante es un tránsito de un planetaide excéntrico? 🙄🤔	fear
txiuuu 🙄	fear

3.9. Despliegue

La fase de Despliegue en la metodología CRISP-DM corresponde a la implementación de los resultados obtenidos en las etapas previas, con el objetivo de generar productos útiles para la toma de decisiones, la comunicación de hallazgos o el soporte a procesos posteriores. Dependiendo del contexto del proyecto, esta etapa puede abarcar desde la entrega de informes hasta la puesta en marcha de sistemas dinámicos de monitoreo o integración con plataformas externas.

En esta memoria, el despliegue se ha enfocado en la generación de visualizaciones de datos que permiten representar gráficamente los resultados del trabajo realizado. Las visualizaciones fueron seleccionadas y diseñadas para comunicar de manera efectiva, y atractiva visualmente, los patrones detectados en el comportamiento y percepción de los usuarios, considerando aspectos como la frecuencia de comentarios, el análisis de sentimientos, la distribución temporal, y la coocurrencia de palabras clave o emojis.

Para la construcción de estas visualizaciones se emplearon principalmente dos herramientas principales:

- Matplotlib: biblioteca de visualización en Python ampliamente utilizada en ciencia de datos, que permitió generar gráficos de barras, distribuciones, y mapas de calor, con un control detallado sobre el diseño, ejes, etiquetas y escalas.
- Flourish: Plataforma en línea de visualización interactiva que facilita la creación de gráficos dinámicos y visualizaciones narrativas. Fue utilizada para representar relaciones entre entidades (por ejemplo, redes de coocurrencia entre emojis o *hashtags*), secuencias temporales y elementos visuales que requieren interacción del usuario.

Un paso clave en el proceso de despliegue fue la identificación de preguntas relevantes que pudieran ser respondidas mediante los datos disponibles, con el objetivo de guiar la selección y diseño de las visualizaciones. Estas preguntas surgieron del análisis exploratorio, y orientaron la construcción de representaciones gráficas que permitieran generar conocimiento útil y comprensible. Las principales interrogantes que guiaron este proceso fueron:

- ¿Cómo varía la percepción de los usuarios en el tiempo?
- ¿Qué temas tienden a generar reacciones predominantemente positivas?
- ¿Qué temas provocan reacciones mayoritariamente negativas?
- ¿Qué temas se asocian a respuestas neutrales o mixtas con connotaciones tanto positivas como negativas?
- ¿Qué temáticas están predominantemente vinculadas a cada emoción específica?

- ¿Qué información de valor puede extraerse a partir del uso de emojis en los comentarios?

A continuación se mostrarán las ilustraciones realizadas para responder cada pregunta.

3.9.1. ¿Cómo varía la percepción de los usuarios en el tiempo?

La Figura 13 ilustra cómo se distribuyen los sentimientos de los comentarios a lo largo del tiempo, utilizándose un color rojo para el sentimiento negativo, color verde para el sentimiento positivo y un gris para el sentimiento neutro. De la visualización se puede observar que en general el sentimiento negativo siempre es menor en comparación a los sentimientos positivos y neutros, lo cual hace posible concluir que en general la percepción de los usuarios ante las publicaciones sobre salud mental, tienen a ser positivas. De hecho, mediante la observación de comentarios clasificados con sentimiento negativo, se puede decir que la clasificación de negativo no necesariamente corresponde a una reacción de detración ante el contenido publicado, sino que también a comentarios que cuentan experiencias difíciles, que nacen a partir del contenido del *post* que invita a que los usuarios expresen estas historias.

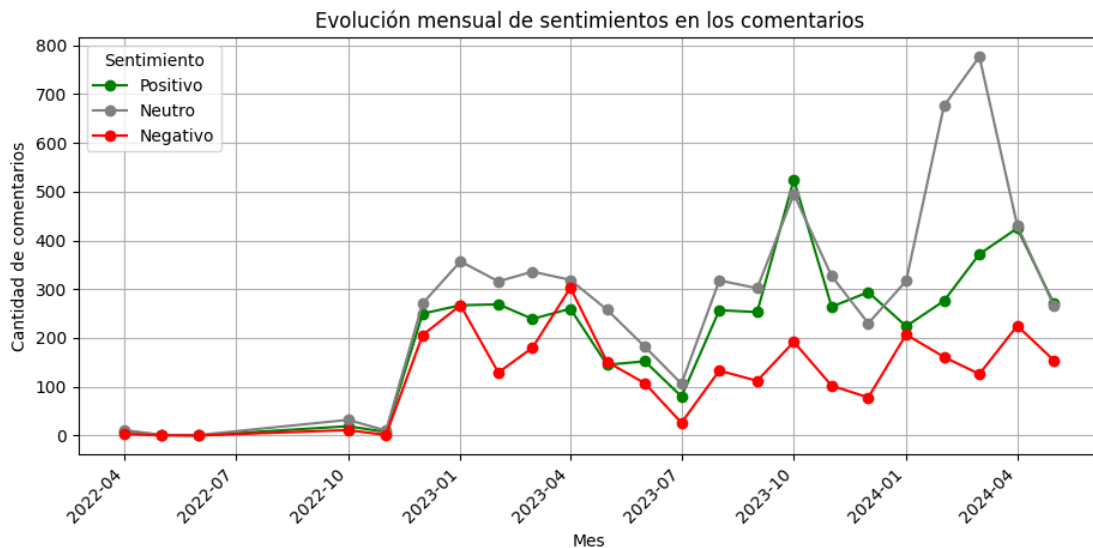


Figura 13: Evolución mensual de los sentimientos en los comentarios, clasificados automáticamente como positivos, neutros o negativos mediante la biblioteca *pysentimiento*.

Fuente: Elaboración propia.

Cabe destacar que los comentarios clasificados como negativos no necesariamente expresan rechazo hacia el contenido publicado. En muchos casos, esta categoría agrupa relatos de experiencias personales difíciles, compartidas a partir del contenido de las publicaciones,

que suelen invitar a la reflexión o a la expresión emocional. A continuación, se presentan algunos ejemplos representativos de este tipo de comentarios:

yo sufro de insomnio, duermo solo cuatro horas y de forma intermitente 😞 si logro dormir siesta me duele la cabeza.

😞 me pasa esto

me pasó, me ha pasado y la verdad no se como hacerlo. no se como salir de esta situación. siento que pierdo todo de mi...

mi trabajo no me lo permite 😞

y si tu trauma justamente impide que te atrevas a establecer nuevas relaciones o ir a lugares nuevos? 😞

En cuanto a los comentarios neutros, se observa que estos son predominantes durante prácticamente todos los meses, con un aumento particularmente notorio en marzo de 2024.

Es importante tener presente que, como se observa en la matriz de confusión del modelo *pysentimiento* (Figura 9), una proporción de los comentarios clasificados como neutros en realidad corresponden a comentarios positivos. Esto implica que la curva asociada al sentimiento neutro en la visualización está ligeramente sobreestimada, mientras que la curva del sentimiento positivo está subestimada, lo que refuerza aún más la conclusión de que la percepción general de los usuarios frente a las publicaciones sobre salud mental es predominantemente positiva.

Además, una parte significativa de los comentarios correctamente clasificados como neutros corresponde a usuarios que recomiendan el contenido a otros mediante el etiquetado de sus nombres de usuario. Este comportamiento también podría interpretarse como una manifestación de percepción positiva, ya que es razonable suponer que al recomendar un contenido, el usuario lo considera valioso o relevante.

Finalmente, la visualización también evidencia una considerable variabilidad en la cantidad de comentarios de cada tipo (negativo, positivo y neutro) a lo largo del tiempo. Si bien explicar con precisión las causas de esta variabilidad escapa al alcance de esta memoria, una posible explicación es que las publicaciones, al tratar temas distintos cada mes, generan respuestas emocionales también variables por parte de los usuarios.

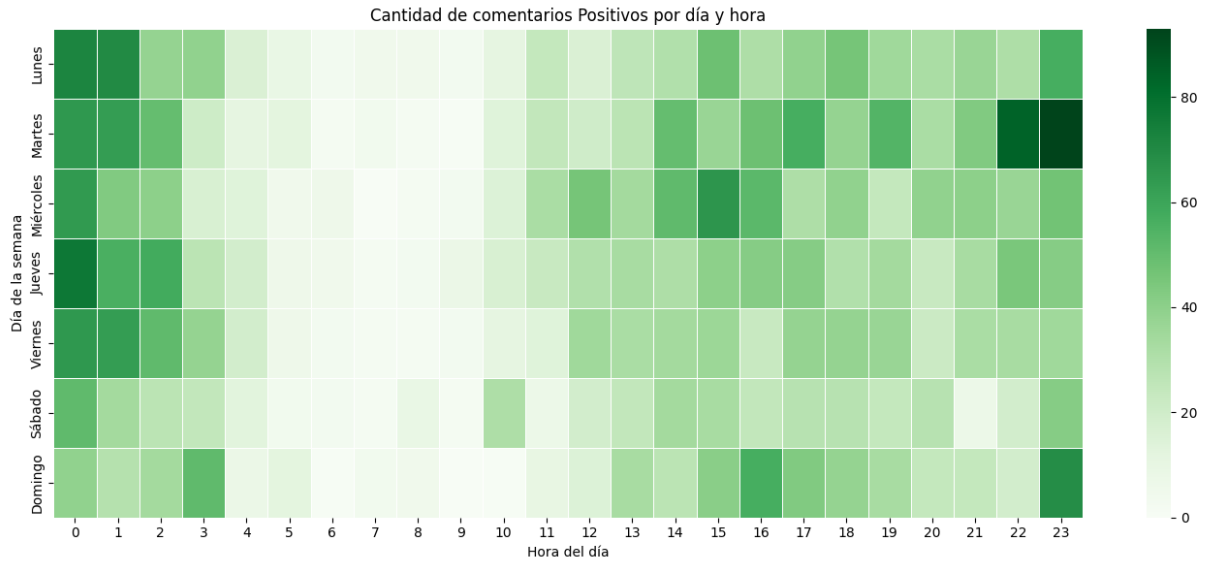


Figura 14: Mapa de calor que representa la cantidad de comentarios con sentimiento positivo, distribuidos por día de la semana y hora del día.

Fuente: Elaboración propia.

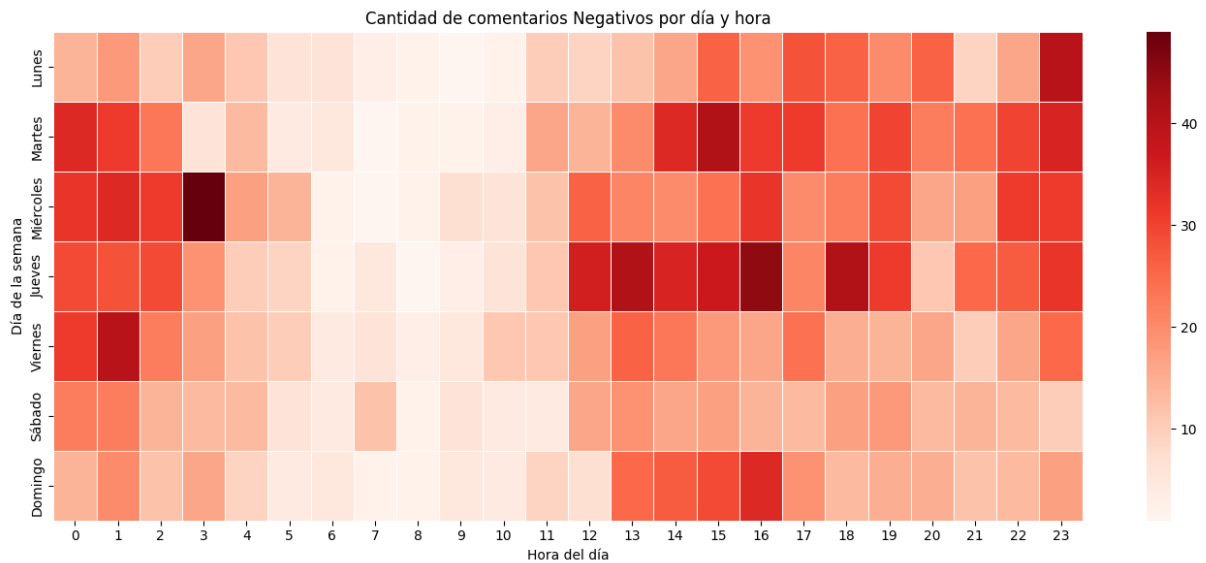


Figura 15: Mapa de calor que representa la cantidad de comentarios con sentimiento negativo, distribuidos por día de la semana y hora del día.

Fuente: Elaboración propia.

Las Figuras 14 y 15 presentan mapas de calor que ilustran la distribución temporal —por día de la semana y hora del día— de los comentarios clasificados con sentimiento positivo y negativo, respectivamente. Estas visualizaciones permiten identificar ciertos patrones que

podrían dar indicios del comportamiento de los usuarios ante publicaciones relacionadas con salud mental.

Uno de los hallazgos más llamativos se observa los domingos por la noche y los lunes de madrugada, donde se produce un aumento significativo en la cantidad de comentarios positivos, mientras que los comentarios negativos en esos horarios son poco frecuentes. Por otro lado, los lunes por la noche destacan por una mayor presencia relativa de comentarios negativos. Aunque el número absoluto de comentarios positivos y negativos en ese horario es similar (aproximadamente 45 en cada caso), el análisis proporcional revela una mayor prevalencia de los negativos. Durante la noche del martes, se aprecia un incremento notorio de comentarios positivos. En contraste, el miércoles a las 3 de la madrugada se registra una concentración inusualmente alta de comentarios negativos, con una escasa presencia de comentarios positivos en ese mismo horario. También en otros momentos específicos de la semana hay alzas de comentarios negativos, tales como el martes a las 15:00 horas, el jueves entre las 13:00 y 18:00 horas, y el viernes a la 1:00 de la madrugada.

Aunque estas visualizaciones permiten identificar ciertos comportamientos temporales que podrían ser relevantes, por sí solas no son suficientes para establecer conclusiones definitivas sobre patrones de conducta. Si bien es plausible pensar que podría existir una tendencia a emitir comentarios negativos durante la madrugada —posiblemente relacionada con el cansancio acumulado, que sería menor durante los fines de semana o inicios de semana, como sugiere la Figura 15, también es posible que estos picos se deban a la publicación de contenido controversial en momentos específicos.

En definitiva, estas visualizaciones constituyen un punto de partida valioso para futuros estudios que busquen profundizar en los patrones temporales de interacción de los usuarios y su relación con los temas abordados en las publicaciones sobre salud mental.

3.9.2. ¿Qué temas tienden a generar reacciones predominantemente positivas?

La Figura 16 ilustra las 15 palabras más frecuentes en publicaciones cuya mayoría de comentarios fueron clasificados con sentimiento positivo, según una regla de mayoría simple. Para este análisis se excluyeron *stopwords* y *hashtags*. En concordancia con visualizaciones anteriores, se utiliza nuevamente el color verde, aunque en una tonalidad distinta para que la visualización sea más agradable visualmente.

Tal como se observa, las palabras más frecuentes son “salud” y “mental”, lo cual resulta esperable considerando que esa es precisamente la temática central de las publicaciones analizadas. En cuanto a otras palabras destacadas, algunas no poseen un significado unívoco. Por ejemplo, la tercera palabra más frecuente, vida, no entrega por sí sola una interpretación clara, pero es razonable suponer que puede aludir a la experiencia vital en general, a la calidad de vida o incluso a una revalorización personal asociada a procesos de salud mental.

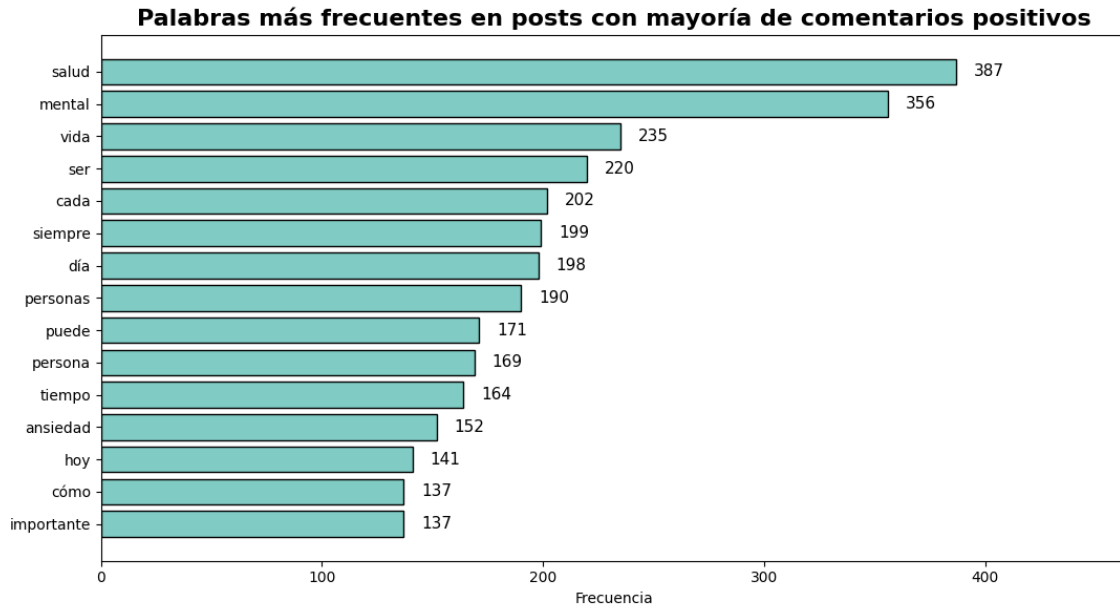


Figura 16: Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios positivos.

Fuente: Elaboración propia.

No obstante, entre las 15 palabras hay una que resulta particularmente interesante: “ansiedad”. La presencia destacada de este término sugiere que las publicaciones que abordan esta temática suelen recibir respuestas mayoritariamente positivas por parte de los usuarios. Esta observación se vincula directamente con el contexto descrito en la investigación sobre estadísticas de salud mental en Chile (véase Sub-sección 3.3.1). En particular, la encuesta Termómetro de la Salud Mental [Bravo *et al.*, 2024], en su octava ronda, indicó que la ansiedad sigue siendo el trastorno más prevalente en la población chilena: un 24,8 % de los encuestados manifestó síntomas de ansiedad, cifra levemente superior al 22,3 % registrado en la medición anterior, consolidando así una tendencia estable desde mediados de 2021.

Dado que la ansiedad es —si no el más prevalente— al menos uno de los trastornos de salud mental más frecuentes en el país, tiene sentido que sea una temática que genera reacciones positivas. Esto puede explicarse porque los usuarios se sienten identificados con el contenido, encuentran valor en la información compartida o se sienten comprendidos emocionalmente al ver representada su experiencia. Así, esta visualización no solo destaca una palabra clave, sino que también evidencia cómo ciertas temáticas logran conectar de manera significativa con la comunidad.

en interacciones entre usuarios, ya sea agradeciendo respuestas positivas o reconociendo el apoyo recibido. Este tipo de dinámica también puede incluir al autor del *post*, quien agradece comentarios favorables o empáticos.

A continuación, se presentan algunos ejemplos reales que reflejan esta tendencia en los comentarios.

muchas gracias por compartir este estudio ❤️ aunque haya sido llevado a cabo en otras partes, es súper importante tener estas estadísticas! aún se vende la idea de “aceptar” o “avalarse” este tipo de actos bajo el manto del empoderamiento femenino. y, si bien, es algo mucho más complejo de analizar para nosotras en latinoamérica dadas las diferencias socio-culturales, es súper significativo entender lo que lleva a los individuos a estas acciones.

es tan importante y a la vez tan complejo... muchas gracias por tu *post*. 🙌

que linda, muchas gracias por estas recomendaciones bella

@user awwe muchas gracias por ese mensajito 🥺❤️

@user gracias pao! tu siempre presente en esta cuenta y apañando el contenido 🔥❤️ tenemos pendientes ah jajajaaj un abrazo huachis 🥰❤️❤️

gracias hernana de mi vida gracias ppr por acompañarme en todo te amo con mi vida ❤️

@user gracias a ud x ser una inspiración 💚

Otras palabras que destacan significativamente en la nube son siempre, excelente, *post*, mejor, encanta, buena, buen y bien. La mayoría de ellas reflejan una forma de elogio o aprobación, mediante las cuales los usuarios expresan su apoyo y valoración positiva hacia el contenido.

En particular, las palabras siempre, vida y ser también aparecen con alta frecuencia, aunque su uso no responde a un único patrón claro. Sin embargo, a partir de un análisis exploratorio, se identificó que la palabra siempre suele emplearse para reforzar ideas de permanencia, constancia y vínculo emocional, como se evidencia en ejemplos de comentarios reales que se presentan a continuación.

gracias como **siempre** por la info mari 💕

@user gracias por estar, por ser mi amiga / hermana , mis más lindos recuerdos están contigo. te llevo conmigo **siempre** y ambas sabemos que pueden pasar mil años y nos se seguiremos queriendo como el día 1. te quiero por **siempre** ❤️ la negra

feliz aniversario @user 🎂. ver crecer un emprendimiento es como ir maternando un hij@ que pasa por múltiples etapas, a veces desafiantes, pero **siempre** reconfortan y sus cumpleaños son lo mejor así que espero disfruten este día de reflexiones en cuanto han sido estos 4 años de trabajo y espero sigan creciendo mucho más porque sus materiales están hechos con mucha dedicación y eso me encanta! espero ganas porque uno no es ninguno y **siempre** l@s psicologas infantiles queremos otro juguete o material para la consulta 🤔

@user ❤️❤️❤️ linda mi katita sabes bien que la agradecida soy yo por tenerte en mi vida ,**siempre** estaré para ti te quiero mucho 🥺❤️

@user gracias pao! tu **siempre** presente en esta cuenta y apañando el contenido 🔥❤️ tenemos pendientes ah jajajaaj un abrazo huachis 🥰❤️

tengo la suerte de tener a mi mejor amiga que **siempre** me consulta y me ayuda, siempre estamos la una para la otra y espero que este *post* ayude a otros vínculos a tener espacios más sanos de contención y escucha ❤️

Naturalmente, realizar un análisis exploratorio detallado para cada palabra no resulta del todo expedito. Sin embargo, este caso particular permite no solo comprender el uso que se le da a la palabra siempre en los comentarios clasificados como positivos, sino también ilustrar el potencial de este tipo de visualizaciones como punto de partida para investigaciones futuras. En ese sentido, sería interesante profundizar en el análisis de las palabras más frecuentes explorando sus usos en distintos contextos semánticos, mediante técnicas como el análisis de concordancia, el modelamiento de tópicos u otras técnicas interesantes para el caso. Estas herramientas permitirían identificar patrones discursivos más complejos y aportar a una comprensión más matizada del lenguaje emocional empleado por los usuarios al interactuar con contenidos relacionados con salud mental.

3.9.4. ¿Qué temas tienden a generar reacciones predominantemente negativas?

Para responder a esta pregunta se realizó la siguiente visualización.

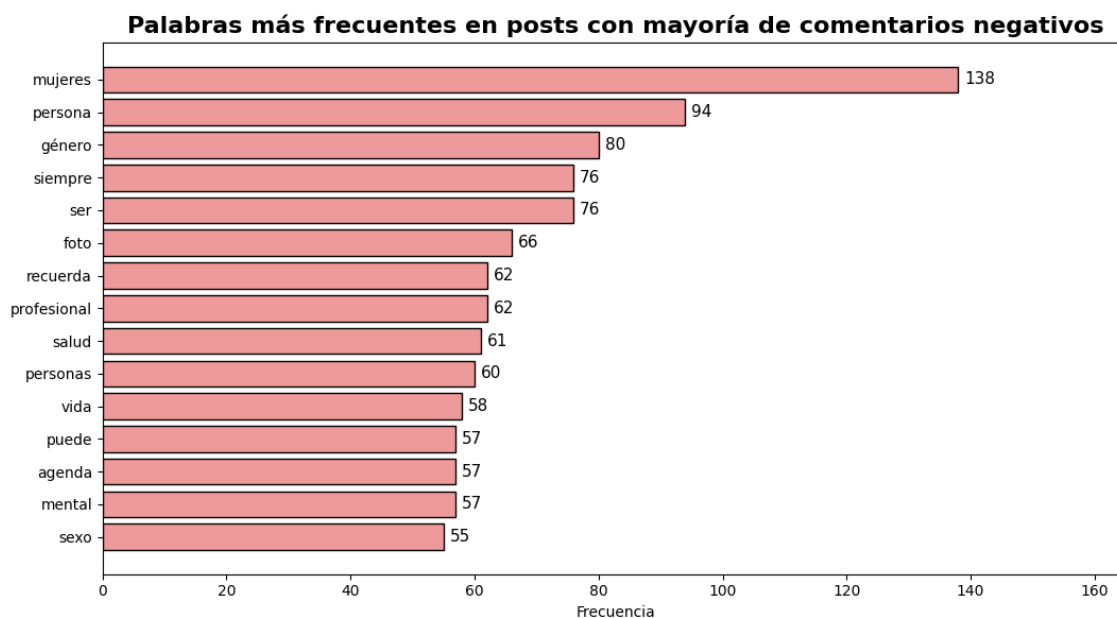


Figura 18: Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios negativos.

Fuente: Elaboración propia.

La Figura 18 ilustra las 15 palabras más frecuentes en publicaciones cuya mayoría de comentarios fueron clasificados con sentimiento negativo, según una regla de mayoría simple. Para este análisis se excluyeron stopwords y *hashtags*. En concordancia con visualizaciones anteriores, se utiliza nuevamente el color rojo, aunque en una tonalidad distinta para que la visualización sea más agradable visualmente.

Al observar la visualización, resulta evidente que varias de las palabras más frecuentes están asociadas a temáticas sociales sensibles en el contexto chileno actual. En particular, términos como “mujeres”, “género” y “sexo” reflejan que muchas de las publicaciones que abordan asuntos relacionados con el feminismo, la diversidad de género o la sexualidad tienden a recibir comentarios cargados de negatividad.

Esta concentración léxica no solo pone en evidencia el carácter controversial que estos temas pueden generar en redes sociales, sino que también invita a reflexionar sobre el clima de polarización que rodea ciertas discusiones públicas, especialmente en espacios digitales donde se expresa libremente la opinión individual.

Más allá de su carga controversial, la presencia de estos temas se relaciona directamente con el contexto descrito en la investigación sobre estadísticas de salud mental en Chi-

A partir de la visualización, se observa con claridad que muchas de las palabras destacadas están asociadas al género, tales como mujeres, ser, mujer, personas y hombres. Esta concentración temática sugiere que una proporción significativa de los comentarios negativos está relacionada con tópicos sensibles como el feminismo y la identidad de género.

Esta observación es especialmente relevante, ya que plantea interrogantes sobre la distribución de los sentimientos negativos: ¿cómo se verían afectadas estas distribuciones si se excluyeran los comentarios asociados a estos temas? Es razonable suponer que la proporción de comentarios negativos podría disminuir si se eliminaran tópicos que tienden a generar mayor polarización o controversia.

Si bien la evaluación cuantitativa de este posible cambio escapa al alcance de esta memoria, abre una línea interesante para investigaciones futuras, orientadas a analizar cómo varían las emociones expresadas en redes sociales según el tema abordado en las publicaciones. Por ejemplo, se podría comparar la distribución de sentimientos en publicaciones que traten salud mental desde una perspectiva clínica, educativa o activista, para observar si ciertos enfoques o temáticas tienden a suscitar reacciones más negativas o positivas por parte de los usuarios.

3.9.6. ¿Qué temáticas están predominantemente vinculadas a la emoción joy?

La Figura 20 ilustra las 15 palabras más frecuentes en publicaciones cuya mayoría de comentarios fueron clasificados con emoción *joy*, según una regla de mayoría simple. Para este análisis se excluyeron *stopwords* y *hashtags*. En concordancia con visualizaciones anteriores, se utiliza nuevamente el color verde, dado que *joy* es la única emoción considerada positiva entre el conjunto de emociones analizadas.



Figura 20: Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios con emoción *joy*.

Fuente: Elaboración propia.

Al observar la visualización —y en relación con su contraparte en el análisis de sentimientos, la Figura 16— se aprecia que una parte importante del vocabulario se mantiene entre ambas representaciones. Esto sugiere una alta correspondencia entre los comentarios clasificados como positivos y aquellos que expresan específicamente alegría. Las diferencias observadas entre ambas visualizaciones probablemente se explican por la aplicación de la heurística basada en emojis utilizada para clasificar emociones (en la sección 3.7.1).

En este sentido, la visualización aporta una capa adicional de interpretación: muchos de los comentarios previamente clasificados con sentimiento positivo pueden, con mayor precisión, entenderse como expresiones de alegría.

3.9.7. ¿Qué temáticas están predominantemente vinculadas a la emoción *anger*?

La Figura 21 ilustra las 15 palabras más frecuentes en publicaciones cuya mayoría de comentarios fueron clasificados con emoción *anger*, según una regla de mayoría simple. Para este análisis se excluyeron *stopwords* y *hashtags*. Se utilizó el color rojo, dada su asociación simbólica con la expresión de la ira.

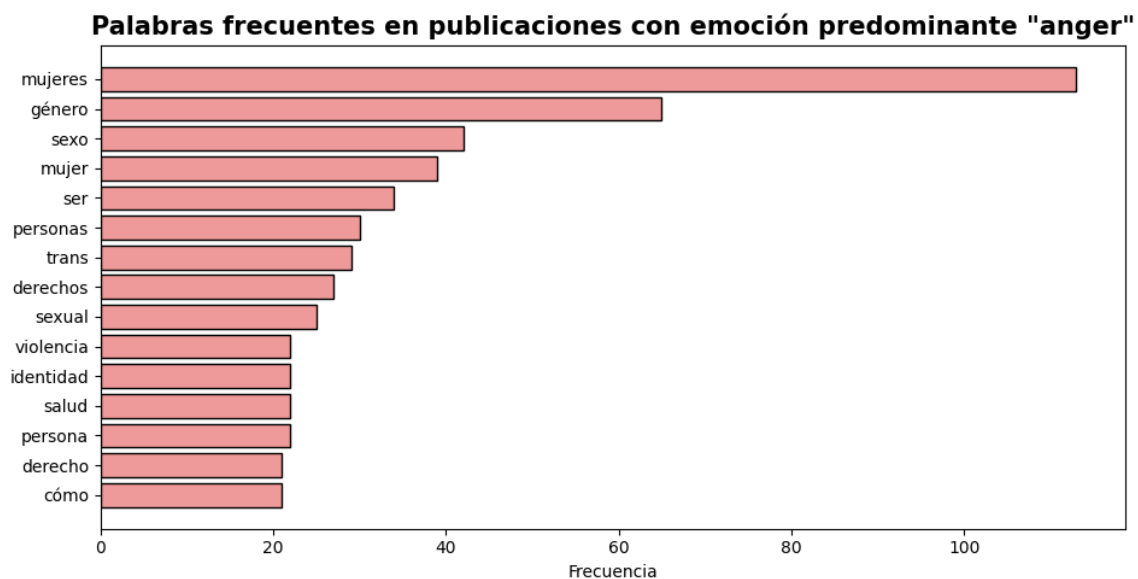


Figura 21: Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios con emoción *anger*.

Fuente: Elaboración propia.

La visualización revela, en consonancia con su análogo en el análisis de sentimientos —la Figura 18—, que los temas relacionados con género y sexualidad, como el feminismo y la identidad de género, siguen siendo focos de controversia en las publicaciones analizadas.

No obstante, esta representación permite ir un paso más allá al mostrar cómo se expresa esa negatividad de forma más específica, en este caso, mediante la emoción del enojo.

Una diferencia relevante respecto a visualizaciones anteriores es la aparición destacada de la palabra “trans”, un término que no figuraba entre las más frecuentes en la clasificación por sentimiento negativo. Esto sugiere que, al menos en lo que respecta a la expresión de la ira, las menciones explícitas a identidades transgénero adquieren un mayor protagonismo. Esta observación abre la posibilidad de que ciertos grupos o temáticas no solo generen mayor volumen de comentarios negativos, sino que además concentren una carga emocional más intensa, como lo es el enojo.

Este hallazgo es relevante porque pone en evidencia cómo algunas discusiones en torno a género y diversidad no solo son controversiales, sino que también despiertan respuestas visiblemente emocionales, lo que podría ser indicativo de resistencias culturales, sesgos o experiencias de conflicto vividas por los usuarios.

3.9.8. ¿Qué temáticas están predominantemente vinculadas a la emoción *sadness*?

La Figura 22 ilustra las 15 palabras más frecuentes en publicaciones cuya mayoría de comentarios fueron clasificados con emoción *sadness*, según una regla de mayoría simple. Para este análisis se excluyeron *stopwords* y *hashtags*. Se utilizó el color azul, dada su asociación simbólica con la expresión de la tristeza.

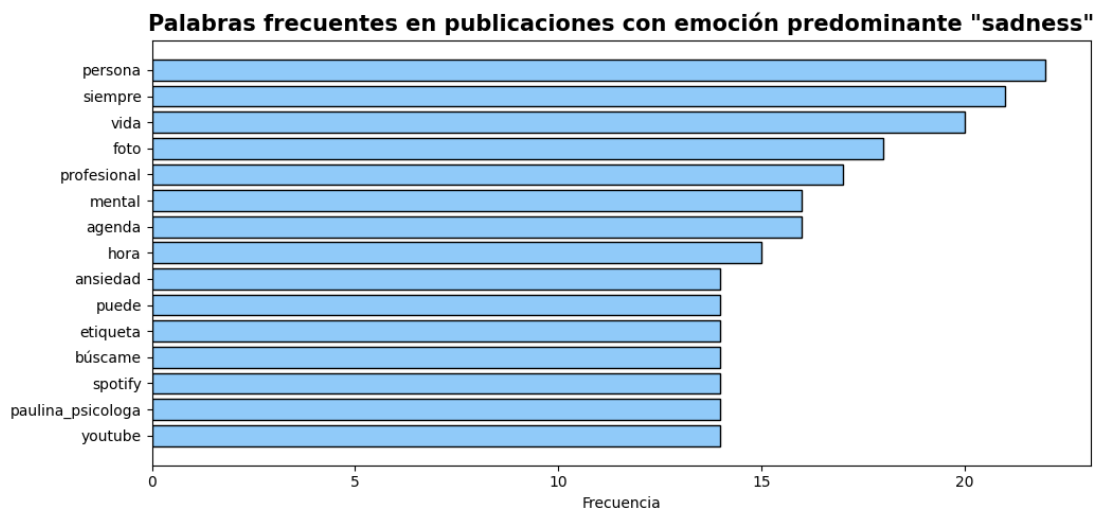


Figura 22: Palabras más frecuentes en publicaciones clasificadas con mayoría de comentarios con emoción *sadness*.

Fuente: Elaboración propia.

De esta visualización es llamativa la aparición de palabras como agenda, hora y nombres propios como paulina_psicologa, además de plataformas como spotify y youtube. Estos términos podrían estar vinculados a contenidos informativos o promocionales relacionados con atención profesional o recursos de apoyo emocional, lo cual sugiere que, incluso en publicaciones que generan o canalizan emociones tristes, los usuarios muestran interés por contenidos útiles, identificables o de carácter terapéutico.

3.9.9. ¿Cómo se expresan en los comentarios cada emoción?

La Figura 23 (la cual se puede observar en mejor calidad en este enlace) presenta una nube de palabras que muestra los términos más utilizados en los comentarios, agrupados según la emoción con la que fueron clasificados. Cada emoción se representa con un color distintivo: Alegría en amarillo, Tristeza en azul, Ira en rojo, Sorpresa en rosado, Asco en verde y Miedo en morado.

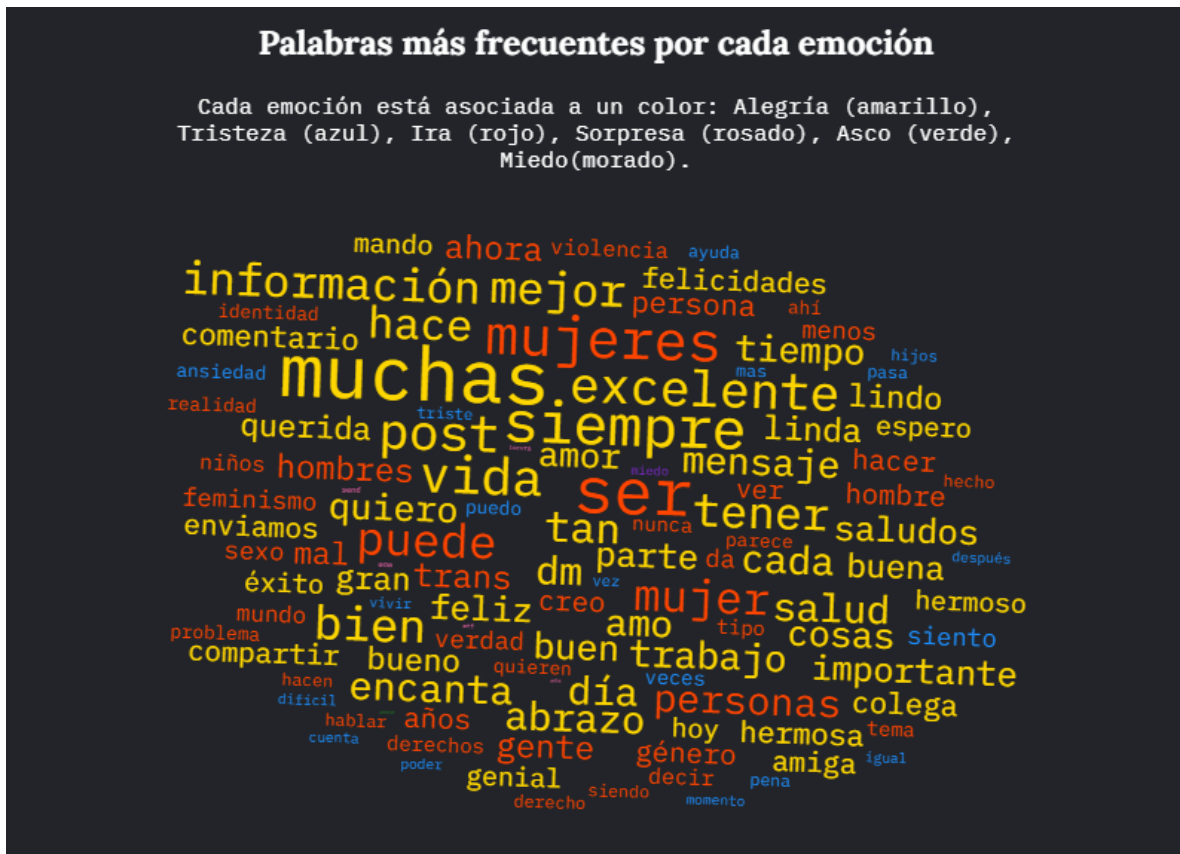


Figura 23: Palabras más frecuentes asociadas a cada emoción.
Fuente: Elaboración propia.

Con el objetivo de generar una visualización más atractiva y comparativa, se optó por integrar todas las emociones en una sola nube de palabras. Esta decisión también responde al hecho de que algunas emociones, al contar con un número reducido de comentarios clasificados, no presentaban un volumen léxico suficiente como para generar una nube representativa de forma individual.

Para asegurar un mínimo de relevancia en los términos visualizados, se incluyó el top 50 de palabras más frecuentes por emoción, aplicando un umbral mínimo de cinco apariciones. Es decir, si una emoción tenía menos de 50 palabras que alcanzaran esa frecuencia, solo se mostraron aquellas que la cumplían. Como consecuencia, ciertos colores aparecen con mayor prominencia que otros, destacando particularmente el amarillo, correspondiente a la emoción Alegría, lo que refleja su mayor presencia en los comentarios.

A partir de la visualización, se puede observar que las emociones que más destacan son Alegría e Ira, lo cual resulta consistente con las Figuras 17 y 19, que mostraban las palabras más frecuentes en comentarios clasificados con sentimientos positivos y negativos, respectivamente. En esta visualización, la Alegría representa la expresión positiva predominante —reflejada por la prominencia del color amarillo—, mientras que entre las emociones negativas, la Ira es la más destacada, seguida por la Tristeza, evidenciada por el tercer color más presente: el azul.

Esta distribución sugiere que una parte importante de los comentarios negativos no necesariamente expresa malestar pasivo o desaprobación general, sino que está cargada de emociones activas como el enojo o la frustración. En segundo lugar, el peso de la Tristeza indica la presencia de comentarios que probablemente expresan dolor emocional, vulnerabilidad o experiencias difíciles. En conjunto, esta visualización no solo refuerza la coherencia léxica observada en las figuras anteriores, sino que además permite distinguir con mayor granularidad el tipo de emoción predominante dentro de cada categoría de sentimiento, aportando así una comprensión más profunda del tono emocional que adoptan los usuarios frente a las publicaciones sobre salud mental.

3.9.10. ¿Qué información de valor puede extraerse a partir del uso de emojis en los comentarios?

La Figura 24 (la cual se puede observar en mejor calidad e interactuar con ella en este enlace) muestra un grafo de coocurrencia de emojis utilizados en los comentarios analizados. En este grafo, los nodos representan emojis individuales, mientras que los arcos indican que dos emojis aparecen juntos en un mismo comentario. El grosor de cada arco corresponde a la frecuencia con la que dicho par de emojis coocurre. Para construir esta visualización se consideraron todos los comentarios disponibles, sin distinción de clasificación emocional o sentimental. Con el fin de mantener la claridad visual y evitar la saturación del grafo, se incluyeron únicamente los 29 emojis más frecuentes.



Figura 24: Grafo de coocurrencia de emojis utilizados en los comentarios.
Fuente: Elaboración propia.

Este grafo actúa como una síntesis visual de algunos hallazgos explorados previamente. Por ejemplo, en el lado derecho de la red se agrupan varios emojis asociados a temáticas de género, como el símbolo femenino, figuras femeninas con expresiones de frustración, y los corazones de color morado y verde, frecuentemente vinculados al movimiento feminista en Chile. Esta agrupación refuerza lo observado en análisis anteriores, donde temas de género tendían a generar comentarios clasificados con emociones negativas, como ira o tristeza.

Por otro lado, una parte considerable del grafo está compuesta por emojis que expresan emociones positivas, como corazones, rostros sonrientes y expresivos de cariño o gratitud, lo cual concuerda con los resultados obtenidos en el análisis de sentimientos y emociones predominantes. Incluso en el caso de emociones más complejas, como la tristeza, es posible identificar emojis como la cara con lágrima o el parche curita, que podrían simbolizar consuelo o vulnerabilidad emocional. En conjunto, esta red permite visualizar de forma condensada las dinámicas emocionales que atraviesan los comentarios, ofreciendo una representación gráfica de las asociaciones afectivas y discursivas que los usuarios construyen mediante el uso de emojis.

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

La validación constituye una etapa crítica dentro del desarrollo de una memoria de título, ya que permite determinar si la solución propuesta cumple con los objetivos planteados y si resulta adecuada para el entorno en el cual se implementaría. Validar implica, en términos generales, someter la propuesta a una evaluación que permita sustentar su pertinencia, utilidad y solidez metodológica, ya sea a través de pruebas técnicas, aplicación en escenarios reales o recolección de opiniones expertas.

En el contexto de este trabajo, enfocado en el análisis de la percepción frente a publicaciones sobre salud mental en una red social en Chile mediante técnicas de *machine learning* (procesamiento de lenguaje natural) y visualización de datos, se optó por realizar una validación cualitativa a través de encuestas dirigidas a profesionales del área de la psicología. Esto, considerando que el juicio experto es una herramienta válida y pertinente cuando se busca evaluar la coherencia y el valor interpretativo de resultados obtenidos.

A continuación, se presenta la metodología utilizada para aplicar la encuesta, los criterios considerados en su diseño, así como un análisis de las respuestas entregadas por dos psicólogas participantes. El objetivo de esta etapa es reflexionar sobre los hallazgos desde una perspectiva profesional y determinar en qué medida la solución propuesta aporta valor al estudio de las percepciones en torno a la salud mental en redes sociales, en Chile.

4.1. Metodología de la encuesta

Para validar la solución propuesta, se diseñó y aplicó una encuesta dirigida a profesionales del área de la psicología, con el objetivo de obtener una evaluación cualitativa experta sobre los resultados obtenidos en este trabajo.

La encuesta fue aplicada a dos psicólogos chilenos con años de experiencia en psicología clínica, quienes fueron invitados a revisar una serie de visualizaciones seleccionadas. Cada visualización fue acompañada de una breve descripción de su contexto y propósito analítico, seguida por una o más preguntas abiertas y/o cerradas orientadas a recoger la opinión profesional. Las preguntas no solo apuntaban a confirmar la validez aparente de los hallazgos, sino también a fomentar una reflexión crítica desde una perspectiva especializada. Dado que los psicólogos cuentan con un conocimiento experto que no estuvo presente durante el análisis técnico realizado en esta memoria, su participación podía aportar observaciones y matices relevantes que enriquecieran la interpretación de los resultados.

La aplicación de la encuesta se realizó de forma remota, mediante el envío de un formulario digital personalizado que incluía un contexto general, las visualizaciones y preguntas corres-

pondientes. Esta modalidad permitió a los participantes responder en sus propios tiempos, favoreciendo respuestas reflexivas y argumentadas.

Con esta metodología, se buscó incorporar una mirada profesional que complementara los análisis cuantitativos y computacionales realizados, aportando así una validación externa desde el campo profesional de la psicología.

4.2. Diseño de la encuesta

La encuesta se dividió en tres secciones. A continuación, se mostrará la estructura de la encuesta y las preguntas incluidas.

4.2.1. Sección 1: Publicaciones (Sentimientos)

En el marco de un estudio sobre la percepción en una red social ante publicaciones de salud mental en Chile, se analizaron publicaciones de Instagram relacionadas con este tema. A cada publicación se le asignó un sentimiento predominante en función de la clasificación de los comentarios que recibió. Por ejemplo, una publicación se considera de emoción positiva cuando la mayoría de los comentarios asociados a ella fueron clasificados automáticamente con ese sentimiento, utilizando un modelo de análisis de sentimientos.

Pregunta 1. La siguiente visualización muestra las palabras más frecuentes en publicaciones clasificadas con el sentimiento predominante "positivo", es decir, aquellas que generaron principalmente respuestas positivas en los usuarios. En base a su experiencia profesional, y considerando que las publicaciones aquí representadas fueron clasificadas como positivas, por contener mayoritariamente comentarios de connotación positiva, ¿considera que las palabras más frecuentes destacadas en esta visualización (por ejemplo: "salud", "mental", "vida", "ser") reflejan temáticas que razonablemente pueden estar asociadas a un sentimiento positivo? Responda Sí, No u Otras en caso de que desee agregar algún comentario.

Para esta pregunta, la visualización mostrada corresponde a la Figura 16.

Pregunta 2. La siguiente visualización muestra las palabras más frecuentes en publicaciones clasificadas con el sentimiento predominante "negativo", es decir, aquellas que generaron principalmente respuestas negativas en los usuarios. En base a su experiencia profesional, y considerando que las publicaciones aquí representadas fueron clasificadas como negativas por contener mayoritariamente comentarios de connotación negativa, ¿considera que las palabras más frecuentes destacadas en esta visualización reflejan temáticas que razonablemente pueden estar asociadas a un sentimiento negativo? Responda Sí, No u Otras en caso de que desee agregar algún comentario.

Para esta pregunta, la visualización mostrada corresponde a la Figura 18.

4.2.2. Sección 2: Publicaciones (Emociones)

En el marco de un estudio sobre la percepción en una red social ante publicaciones de salud mental en Chile, se analizaron publicaciones de Instagram relacionadas con temas de salud mental en Chile. A cada publicación se le asignó una emoción predominante en función de la clasificación de los comentarios que recibió. Por ejemplo, una publicación se considera de emoción "joy" (alegría) cuando la mayoría de los comentarios asociados a ella fueron clasificados automáticamente con esa emoción, utilizando un modelo de análisis de emociones.

Pregunta 3. La siguiente visualización muestra las palabras más frecuentes en publicaciones clasificadas con la emoción predominante "joy", es decir, aquellas que generaron principalmente respuestas de alegría en los usuarios. En base a su experiencia profesional, y considerando que las publicaciones aquí representadas fueron clasificadas como alegres por contener mayoritariamente comentarios asociados a la emoción "joy", ¿considera que las palabras más frecuentes destacadas en esta visualización (por ejemplo: "salud", "mental", "vida", "ser") reflejan temáticas que razonablemente pueden estar asociadas a emociones de alegría o bienestar emocional? Responda Sí, No u Otras en caso de que desee agregar algún comentario.

Para esta pregunta, la visualización mostrada corresponde a la Figura 20.

Pregunta 4. La siguiente visualización muestra las palabras más frecuentes en publicaciones clasificadas con la emoción predominante *anger*, es decir, aquellas que generaron principalmente respuestas de ira o enojo en los usuarios. En base a su experiencia profesional, y considerando que las publicaciones aquí representadas fueron clasificadas como de enojo por contener mayoritariamente comentarios asociados a la emoción *anger*, ¿considera que las palabras más frecuentes destacadas en esta visualización reflejan temáticas que razonablemente pueden estar asociadas a emociones de ira o molestia? Responda Sí, No u Otras en caso de que desee agregar algún comentario.

Para esta pregunta, la visualización mostrada corresponde a la Figura 21.

Pregunta 5. La siguiente visualización muestra las palabras más frecuentes en publicaciones clasificadas con la emoción predominante "sadness", es decir, aquellas que generaron principalmente respuestas de tristeza en los usuarios. En base a su experiencia profesional, y considerando que las publicaciones aquí representadas fueron clasificadas como tristes por contener mayoritariamente comentarios asociados a la emoción "sadness", ¿considera que las palabras más frecuentes destacadas en esta visualización reflejan temáticas que razonablemente pueden estar asociadas a emociones de tristeza o melancolía? Responda Sí, No u

Otras en caso de que desee agregar algún comentario.

Para esta pregunta, la visualización mostrada corresponde a la Figura 22.

4.2.3. Sección 3: Comentarios

Pregunta 6. La siguiente visualización muestra la cantidad de comentarios con carga emocional negativa según el día de la semana y la hora del día en publicaciones de Instagram relacionadas con salud mental en Chile. Esta información fue obtenida a partir de un modelo automatizado de análisis de sentimientos aplicado a los comentarios. En base a su experiencia profesional: ¿Considera que existen momentos del día o de la semana donde es más probable que las personas expresen emociones negativas en redes sociales, y que esta visualización refleja razonablemente esa tendencia? Comentarios adicionales o interpretaciones pueden ser agregados en Otras.

Opciones:

- Sí, considero que la visualización refleja patrones que tienen sentido desde una perspectiva clínica o social.
- Parcialmente, algunos momentos parecen consistentes pero otros no tanto.
- No, no considero que la visualización represente un patrón significativo.
- Otras (con la opción de escribir).

Para esta pregunta, la visualización mostrada corresponde a la Figura 15.

Pregunta 7. La siguiente visualización muestra la cantidad de comentarios con carga emocional positiva según el día de la semana y la hora del día en publicaciones de Instagram relacionadas con salud mental en Chile. Esta información fue obtenida a partir de un modelo automatizado de análisis de sentimientos aplicado a los comentarios. En base a su experiencia profesional: ¿Considera que existen momentos del día o de la semana donde es más probable que las personas expresen emociones positivas en redes sociales, y que esta visualización refleja razonablemente esa tendencia? Comentarios adicionales o interpretaciones pueden ser agregados en Otras.

Opciones:

- Sí, considero que la visualización refleja patrones que tienen sentido desde una perspectiva clínica o social.
- Parcialmente, algunos momentos parecen consistentes pero otros no tanto.

- No, no considero que la visualización represente un patrón significativo.
- Otras (con la opción de escribir).

Para esta pregunta, la visualización mostrada corresponde a la Figura 14.

Pregunta 8. La siguiente visualización presenta una nube de palabras generada a partir de los términos más frecuentes en comentarios con sentimiento positivo en publicaciones de Instagram relacionadas con temas de salud mental. Las palabras de mayor tamaño corresponden a aquellas que se repiten con mayor frecuencia. ¿Qué tipo de temas, emociones o dinámicas relacionales considera usted que se reflejan en esta visualización? ¿Le parece que esta representación entrega información útil o relevante para comprender cómo las personas se vinculan con los contenidos de salud mental desde una perspectiva emocional positiva? Puede visualizar mejor e interactuar con esta visualización en: <https://public.flourish.studio/visualisation/23655906/>

Para esta pregunta, la visualización mostrada corresponde a la Figura 17.

Pregunta 9. La siguiente visualización presenta una nube de palabras generada a partir de los términos más frecuentes en comentarios con sentimiento negativo en publicaciones de Instagram relacionadas con temas de salud mental. Las palabras de mayor tamaño corresponden a aquellas que se repiten con mayor frecuencia. ¿Qué tipo de malestares, preocupaciones o experiencias cree usted que se manifiestan a través de estas palabras? ¿Considera que esta representación permite acceder a una comprensión significativa del modo en que las personas expresan emociones negativas en el contexto de redes sociales? Puede visualizar mejor e interactuar con esta visualización en: <https://public.flourish.studio/visualisation/23656135/>

Para esta pregunta, la visualización mostrada corresponde a la Figura 19.

Pregunta 10. La siguiente visualización muestra una nube de palabras donde cada término está asociado a una emoción predominante mediante el uso del color. ¿Qué impresiones o interpretaciones le genera esta representación? ¿Le parece que entrega información útil o significativa sobre el lenguaje emocional expresado en redes sociales? Puede visualizar mejor e interactuar con esta visualización en: <https://public.flourish.studio/visualisation/23656386/>

Para esta pregunta, la visualización mostrada corresponde a la Figura 23.

Pregunta 11. La siguiente visualización representa una red de coocurrencia de emojis en comentarios de publicaciones sobre salud mental en Instagram. Cada nodo representa un emoji, y los enlaces entre ellos indican que fueron utilizados juntos en el mismo comentario.

La intensidad del vínculo refleja cuán frecuente es esa coocurrencia. Desde su experiencia profesional: ¿Considera que este tipo de representación gráfica aporta información relevante o interesante sobre la manera en que las personas comunican emociones en redes sociales mediante emojis? ¿Qué tipo de relaciones emocionales, significados compartidos o patrones comunicativos cree que podrían interpretarse a partir de este tipo de visualización? Puede visualizar mejor e interactuar con esta visualización en: <https://public.flourish.studio/visualisation/23656614/>

Para esta pregunta, la visualización mostrada corresponde a la Figura 24.

4.2.4. Sección 4: Preguntas generales y de perfil

Pregunta 12. Desde una perspectiva psicológica, y considerando el impacto psicosocial que pueden tener diversas problemáticas contemporáneas, ¿en qué medida considera que temáticas como el feminismo, la diversidad de género (incluyendo experiencias trans), el maltrato animal, y aspectos vinculados a la educación y al contexto escolar, pueden ser abordadas como dimensiones relevantes dentro del ámbito de la salud mental, particularmente en el contexto de su expresión y recepción en redes sociales?

Opciones:

- Totalmente de acuerdo.
- De acuerdo.
- Ni de acuerdo ni en desacuerdo.
- En desacuerdo.
- Totalmente en desacuerdo.
- Otras (con opción de escribir).

Pregunta 13. ¿Cuál es su nombre?

Pregunta 14. ¿Cuántos años de experiencia tiene ejerciendo como psicólogo/a?

Pregunta 15. ¿Con qué sexo se identifica?

Pregunta 16. ¿Cuántos años tiene?

Pregunta 17. ¿En qué área(s) de la psicología se desempeña actualmente?

4.3. Análisis de las respuestas

4.3.1. Con respecto a las visualizaciones de palabras más frecuentes.

Las respuestas entregadas por ambos psicólogos coinciden en que las palabras destacadas en las visualizaciones permiten identificar temáticas que efectivamente pueden generar una determinada carga emocional en los usuarios. En este sentido, las visualizaciones no solo reflejan palabras frecuentes, sino también ofrecen una aproximación al tipo de contenidos que provocan respuestas emocionales específicas en el contexto de redes sociales.

Desde una perspectiva clínica, se valoró que muchas de estas palabras remiten a preocupaciones cotidianas y existenciales, como el bienestar, la identidad personal o el rol social, todas dimensiones estrechamente ligadas a la salud mental. Por ejemplo, términos como salud o vida fueron interpretados como representaciones de búsqueda de bienestar, superación o autocuidado, aspectos que pueden explicar su presencia en publicaciones con predominancia de sentimientos positivos.

Sin embargo, también se destacó que ciertas palabras —como ansiedad, tiempo o profesional— pueden evocar reacciones emocionales ambivalentes. Uno de los encuestados observó que estos términos, aunque presentes en contextos positivos, podrían igualmente despertar malestar dependiendo del significado que cada persona les atribuya. Así, tiempo podría vincularse tanto a la organización personal y la mejora del estilo de vida, como a la presión social, el agotamiento o la falta de equilibrio.

Por otro lado, se problematizó la aparición de términos como mujeres, género o sexo en publicaciones clasificadas como negativas, sugiriendo que estas temáticas —asociadas a cuestiones de identidad, derechos o equidad— pueden generar controversia o incomodidad en ciertos sectores sociales. Este punto fue interpretado como un reflejo de tensiones socio-culturales aún presentes, donde discursos vinculados a feminismo o diversidad de género pueden despertar reacciones emocionales intensas, ya sea de apoyo o rechazo.

En síntesis, los profesionales coincidieron en que las palabras destacadas en las visualizaciones entregan información significativa sobre los temas que gatillan respuestas emocionales específicas. Si bien reconocieron que la interpretación del lenguaje es compleja y situada, también valoraron que estas representaciones permiten detectar patrones discursivos y emocionales que resultan pertinentes desde el punto de vista clínico y psicosocial.

4.3.2. Sobre las visualizaciones temporales.

En relación con las visualizaciones sobre la expresión de emociones según días y horarios, uno de los psicólogos consideró que estas no mostraban patrones significativos, mientras que el otro opinó que sí reflejaban tendencias observables en el quehacer clínico, como la

expresión de emociones negativas al final del día o a mitad de semana, asociadas a la fatiga y el estrés acumulado.

Estas diferencias sugieren que la validez de los patrones temporales puede depender del marco teórico o clínico que cada profesional utilice, lo cual es coherente con la diversidad de enfoques existentes en la psicología.

4.3.3. Sobre las visualizaciones de nubes de palabras.

Ambos participantes valoraron positivamente las nubes de palabras como herramientas útiles para identificar temas y dinámicas relacionales presentes en los comentarios. Se destacaron palabras asociadas al agradecimiento, la esperanza, el acompañamiento y la superación como señales de una expresión emocional positiva en el contexto de salud mental.

En el caso de las palabras negativas, se interpretaron como manifestaciones de malestar, rechazo social, y experiencias difíciles como las vividas por personas trans o en contextos de exclusión. La nubes de emociones con codificación por color fue particularmente bien valorada, y se destacó su potencial para representar de manera sintética la variedad emocional expresada por los usuarios. También se comentó que los términos asociados a emociones menos frecuentes parecen ser menos relevantes, lo cual coincide con lo que la visualización buscaba transmitir.

4.3.4. Sobre la visualización de coocurrencia de emojis.

Ambos encuestados consideraron que la representación de coocurrencia de emojis aporta información interesante sobre la forma en que las personas comunican emociones de manera no verbal en redes sociales. Se señaló que los emojis permiten una expresión emocional sintética y que reflejan patrones actuales de comunicación afectiva. Además, se valoró positivamente que esta visualización permita captar significados compartidos que trascienden las palabras.

4.3.5. Sobre la relación entre temáticas sociales y salud mental.

En la pregunta general sobre la pertinencia de abordar temas como el feminismo, la diversidad de género, el maltrato animal y la educación en el marco de la salud mental, ambos participantes se manifestaron totalmente de acuerdo. Destacaron que todas las experiencias personales y sociales tienen un impacto emocional y que, por tanto, deben ser consideradas dentro del campo de la salud mental, especialmente cuando se expresan en redes sociales.

4.3.6. Sobre quienes respondieron.

Fueron dos psicólogos quienes amablemente respondieron esta encuesta:

- Tania, con más de 10 años de experiencia como psicóloga, Mujer, entre 36 y 45 años y actualmente desempeñándose en la psicología clínica.
- Ignacio, entre 4 y 6 años de experiencia como psicólogo, Hombre, entre 26-35 años y también desempeñándose en la psicología clínica.

Ambos psicólogos dedicaron aproximadamente 1 hora a responder esta encuesta.

4.4. Conclusión de la Validación

A partir del análisis de las respuestas entregadas por los psicólogos encuestados, es posible concluir que las visualizaciones generadas lograron cumplir con su objetivo central: reflejar la percepción de los usuarios ante publicaciones sobre salud mental en una red social en Chile.

Uno de los aspectos más valiosos de la validación fue constatar que las visualizaciones ofrecieron a los profesionales un punto de partida legítimo para emitir interpretaciones desde su experiencia. Cada psicólogo, desde su formación y trayectoria, pudo identificar temáticas, dinámicas emocionales y patrones discursivos que emergen del lenguaje utilizado en los comentarios analizados. Esto no solo da cuenta de la utilidad del enfoque propuesto, sino que también refuerza su capacidad para generar insumos interpretativos válidos desde una perspectiva psicológica.

Si bien algunas visualizaciones fueron consideradas más significativas o claras que otras, todas generaron reflexión y permitieron extraer observaciones relevantes. En este sentido, la validación no solo confirma la coherencia general del análisis, sino que también destaca el valor de combinar métodos computacionales con la interpretación profesional experta para enriquecer la comprensión de fenómenos complejos como la expresión emocional en entornos digitales.

Finalmente, la participación de los psicólogos permitió confirmar que los resultados no solo poseen validez técnica, sino también pertinencia clínica y social. Y dado que las visualizaciones lograron cumplir su objetivo principal, se puede afirmar que el proceso analítico y metodológico que las sustenta también alcanza un grado de validez, en tanto permitió generar representaciones comprensibles, pertinentes y significativas desde una perspectiva profesional.

CONCLUSIONES

Alcance

El propósito central de este estudio fue conocer cómo perciben los usuarios de una red social en Chile —en este caso, Instagram— el contenido publicado en relación con la salud mental. Para ello, fue indispensable cumplir con una serie de pasos esenciales, sin los cuales el desarrollo del trabajo no habría sido posible. Uno de los primeros y más fundamentales consistió en la obtención de los datos necesarios para el análisis, ya que no existía un conjunto de datos público que permitiera realizar el estudio planteado. Para resolver este obstáculo, se implementó con éxito un sistema híbrido compuesto por un *script* propio y una API comercial. A través de esta última, se obtuvo una cantidad aceptable de publicaciones que contenían *hashtags* relacionados con temas de salud mental en Chile, mientras que el *script* desarrollado permitió el *scraping* de todos los comentarios asociados a dichas publicaciones. Si bien fue un desafío comprender el funcionamiento de Instagram para lograr estos objetivos, se consiguió con éxito.

Otro paso crucial fue la recopilación de información relevante sobre la salud mental en Chile, tarea que se abordó mediante la búsqueda y análisis de estadísticas provenientes de diversas fuentes confiables. Esto permitió construir una base sólida de conocimiento, que sirvió como guía para el trabajo posterior. Aunque siempre es posible profundizar más en la recopilación de información, el *background* generado se considera adecuado para los alcances de este estudio, cumpliendo una función orientadora tanto directa como indirectamente en el cumplimiento del objetivo principal.

El trabajo con los datos también resultó exitoso, ya que, a partir de una gran cantidad de información disponible, se logró seleccionar aquellos datos que finalmente aportaron valor significativo al análisis.

La elección de los modelos de análisis también fue adecuada. Aunque uno de los modelos utilizados —*Google Natural Language*— no ofreció el rendimiento esperado, sí permitió establecer un punto de comparación con el modelo principal empleado, *pysentimiento*. Este último entregó buenos resultados, aunque presentó una tendencia a clasificar como neutros ciertos comentarios que claramente eran positivos. Esta limitación, sin embargo, no afectó de manera considerable la calidad del análisis. De hecho, el uso de ambos modelos enriqueció el estudio, ya que permitió contrastar sus resultados y reforzar la elección de *pysentimiento* como herramienta principal. Esta comparación evidenció cómo distintas herramientas pueden diferir en la interpretación del lenguaje y subrayó la importancia de evaluar críticamente sus salidas para obtener resultados más fiables.

La aplicación de la metodología también fue acertada. En particular, la elección de la metodología CRISP-DM resultó ser adecuada, dado su enfoque iterativo. Esta característica fue especialmente valiosa en este proyecto, que se caracterizó por requerir descubrimientos progresivos a medida que avanzaba el análisis, lo que hizo que la metodología calzara como anillo

al dedo.

Por último, la parte más visual del estudio — las visualizaciones de datos — también se ejecutó con éxito. Se logró transformar los datos recolectados en información comprensible y valiosa, que fue evaluada por dos psicólogos. A partir de sus conocimientos especializados, ambos pudieron extraer conclusiones significativas. Esta validación externa respalda el valor de la información obtenida, ya que ofrece una perspectiva adicional, desde un enfoque acotado a los alcances de esta memoria, sobre la situación de la salud mental en Chile.

Limitaciones

Este trabajo presentó importantes limitaciones que, si bien no impidieron alcanzar el objetivo principal, son relevantes de considerar. La primera de ellas está relacionada con el acceso a los datos, ya que no existe una solución oficial para obtener información desde Instagram. Además, una de las herramientas *open source* más conocidas para este fin, Instaloader, dejó de ser útil para los requerimientos específicos de esta investigación. Por otro lado, las APIs comerciales de buena calidad para acceder a la información deseada son, en general, costosas o no ofrecen exactamente las funcionalidades necesarias. Por esta razón, la cantidad de datos obtenidos tuvo que ajustarse a lo permitido por la versión gratuita de la API comercial utilizada. Si se dispone de presupuesto, esta restricción puede mitigarse, pero sigue siendo un aspecto esencial a tener en cuenta al planificar estudios similares.

Otra limitación, también derivada de las dificultades para acceder a los datos de Instagram, fue la necesidad de trabajar exclusivamente con publicaciones que contenían *hashtags*. Esta condición introduce un nivel adicional de ruido en los datos, ya que no se puede asegurar con certeza que todas las personas que comentaron en dichas publicaciones residan en Chile. Nuevamente, con recursos económicos disponibles, sería posible invertir en realizar solicitudes adicionales con el fin de obtener información sobre la ubicación de las publicaciones, lo que permitiría una segmentación más precisa del público objetivo.

Otras limitaciones naturales provienen del carácter automatizado de ciertas etapas del trabajo, especialmente en el análisis de sentimientos. En este caso, la clasificación no fue realizada por expertos humanos en las áreas pertinentes, sino por modelos automatizados. Esto introduce inevitablemente un margen de error o ruido en la información generada, dado que los algoritmos no siempre logran interpretar con precisión la complejidad del lenguaje humano y sus matices emocionales.

Resultados del experimento realizado

Los resultados de este trabajo consistieron en la elaboración de un *dataset* compuesto por publicaciones y comentarios de Instagram relacionados con la salud mental en Chile. Estos datos fueron clasificados en términos de sentimientos (positivos, negativos y neutros) y emociones (alegría, ira, tristeza, asco, sorpresa, miedo y otras). Gracias a este conjunto de datos, fue posible generar visualizaciones que permitieron extraer la siguiente información:

En Chile, la opinión que manifiestan los usuarios frente a publicaciones sobre salud mental es mayoritariamente positiva. Esta conclusión se apoya en el análisis de los comentarios clasificados como neutros, de los cuales una parte efectivamente corresponde a expresiones neutrales. Sin embargo, al observar el comportamiento de los usuarios —por ejemplo, cuando recomiendan el contenido a otras personas— puede interpretarse que estos comentarios reflejan una actitud positiva, incluso si el modelo los clasificó como neutros. Además, se detectó que uno de los modelos utilizados tendió a confundir ciertos comentarios positivos con neutros, lo que refuerza la interpretación de que el sentimiento predominante es positivo.

Dentro de las interacciones positivas, una gran mayoría —si no todas— están asociadas a expresiones de alegría. Esta alegría parece derivar de múltiples factores, como sentirse identificado, comprendido, validado o acompañado por el contenido de las publicaciones. Cabe destacar que gran parte de este contenido es informativo, con recomendaciones y mensajes de apoyo útiles para personas que enfrentan problemas de salud mental. Estos recursos, al abordar de forma empática y clara temas sensibles, permiten que los usuarios se sientan escuchados o representados, lo cual genera reacciones emocionalmente positivas. La alegría, por tanto, no surge solo del contenido en sí, sino de la conexión emocional que los usuarios logran establecer con él.

Otro hallazgo relevante fue que muchas publicaciones etiquetadas bajo temas de salud mental incluyen contenido que, si bien no aborda explícitamente temas de salud mental, sí resulta pertinente pues son temas importantes en la sociedad chilena actual, que generan controversia y por ende, afectan a la salud de la población. Por ejemplo, el feminismo es un tema recurrente: muchas mujeres se ven afectadas por experiencias machistas. De igual forma, la identidad de género fue un tema notable y tiene sentido pues personas no cisgénero enfrentan desafíos que pueden afectar gravemente su salud mental, tales como la discriminación, la disforia o la falta de representación en los medios y discursos públicos. Estos temas —entre otros— generan un fuerte impacto emocional en quienes los viven, por lo que es comprensible que aparezcan recurrentemente en publicaciones relacionadas con salud mental.

Una parte importante de las reacciones negativas observadas en torno a contenido sobre salud mental corresponde a emociones de enojo, las cuales se relacionan directamente con los grupos afectados por las problemáticas antes mencionadas. Estas reacciones no reflejan un rechazo al contenido que aborda directamente temas de salud mental (como, por ejemplo, recomendaciones para manejar la ansiedad), sino una respuesta emocional frente a situaciones que afectan la salud mental de los usuarios de manera estructural o cotidiana.

Asimismo, otra proporción significativa de las reacciones negativas no expresa rechazo, sino tristeza. Esta emoción aparece con frecuencia en comentarios donde los usuarios comparten experiencias personales dolorosas al sentirse identificados con lo planteado en las publicaciones. En estos casos, el contenido funciona como un catalizador que permite a los usuarios verbalizar sentimientos reprimidos o poco visibilizados, lo que otorga un valor adicional a las publicaciones como espacios de expresión y catarsis emocional.

Aplicaciones

Este trabajo contribuye a comprender la percepción que tienen los usuarios de Instagram sobre las publicaciones relacionadas con la salud mental en Chile. Si bien este conocimiento está acotado por los límites propios de una memoria de investigación, puede servir como base para estudios a mayor escala, con aplicaciones concretas en la formulación de políticas públicas. Las redes sociales permiten observar aspectos que no siempre se reflejan en instrumentos tradicionales como las encuestas o los test estandarizados: se captan expresiones auténticas de las personas, los temas que generan interés, así como aquellos que afectan con mayor intensidad la salud mental de la población.

Esta información es especialmente valiosa porque surge de forma espontánea y en contextos naturales de interacción, lo que le otorga un alto grado de autenticidad. No obstante, es importante considerar que las conclusiones se aplican principalmente al perfil etario de los usuarios activos de Instagram, por lo que los resultados pueden no representar adecuadamente a niños pequeños o adultos mayores. Aun así, el análisis de este tipo de datos puede ser útil para detectar prejuicios, patrones de pensamiento y emociones predominantes, elementos clave para diseñar campañas de prevención más efectivas y con enfoques adaptados a la realidad del público objetivo.

Más allá del ámbito de las políticas públicas, este trabajo también aporta al campo del análisis social y de salud desde datos no estructurados. Permite aterrizar la idea de que, a través de publicaciones y comentarios en redes sociales, es posible obtener información relevante sobre la salud mental de las personas en Chile. Aunque el objetivo principal de este estudio no fue evaluar directamente el estado de salud mental de los usuarios, el hecho de conocer su percepción sobre estos temas representa un paso significativo en esa dirección. Se abre así la posibilidad de complementar otras fuentes de información tradicionales con datos derivados de interacciones digitales.

Sin duda, la aplicación de este enfoque a una escala mayor —sin las restricciones de tiempo que impone una memoria académica— podría generar un impacto social relevante, brindando recursos valiosos para mejorar el bienestar psicológico de la población chilena.

Trabajo Futuro

Algunas posibles líneas de trabajo futuro incluyen:

Realizar un estudio similar en plataformas como TikTok o los Reels de YouTube, lo cual implicaría construir herramientas específicas para llevar a cabo el *scraping* de datos en dichos entornos. TikTok, en particular, al ser una red social con un nivel de uso significativamente mayor que Instagram, podría proporcionar información más diversa, abarcando a una mayor variedad de personas y opiniones dentro de la población chilena. Si no existieran restricciones de tiempo, recolectar datos diariamente durante varios meses permitiría obtener un conjunto de datos robusto y valioso, con un potencial analítico aún mayor que el alcanzado en este trabajo.

Otra proyección interesante sería incorporar el análisis de las imágenes contenidas en las

publicaciones. Extraer información visual —y combinarla con los textos de las descripciones y los comentarios— permitiría realizar un análisis más profundo mediante técnicas como el modelado de tópicos. De esta forma, sería posible identificar no solo los *hashtags* o palabras clave, sino también los temas subyacentes que se abordan de manera directa o implícita en las redes sociales al tratar cuestiones relacionadas con la salud mental.

Estas líneas de trabajo permitirían ampliar significativamente el alcance y la profundidad del análisis, aportando nuevas perspectivas sobre cómo se manifiestan y perciben los temas de salud mental en las redes sociales.

REFERENCIAS BIBLIOGRÁFICAS

- [Asociación Chilena de Seguridad, 2025] Asociación Chilena de Seguridad (2025). Una de cada siete personas en Chile presentan síntomas moderados o severos de depresión. <https://www.achs.cl/centro-de-noticias/noticia/2025/una-de-cada-siete-personas-en-chile-presentan-s%C3%ADntomas-moderados-o-severos-de-depresi%C3%B3n>. Noticia.
- [Bravo *et al.*, 2024] Bravo, D., Errázuriz, A., Calfucoy, P., y Campos, D. (2024). Termómetro de la salud mental en Chile ACHS-UC: Octava ronda. Presentación, Centro de Estudios y Encuestas Longitudinales UC y Asociación Chilena de Seguridad (ACHS). Octava ronda del estudio ACHS-UC.
- [Centro de Investigación Periodística, 2021] Centro de Investigación Periodística (2021). Salud mental en Chile: Urgencias, desafíos y silencios. <https://www.ciperchile.cl/2021/11/08/salud-mental-en-chile-urgencias-desafios-y-silencios/>. Visitado el 8 de octubre de 2023, 16:45.
- [DECIDECHILE, 2023] DECIDECHILE (2023). Consejo constitucional. <https://live.decidechile.cl/2023/consejeros>. Visitado el 13 de noviembre de 2023, 20:00.
- [Dewi y Chandra, 2019] Dewi, L. C. y Chandra, A. (2019). Social media web scraping using social media developers api and regex. *Procedia Computer Science*, 157:444-449.
- [Di Bella, 2025] Di Bella, L. (2025). Introducción al análisis de sentimientos en Python.
- [Dykes, 2016] Dykes, B. (2016). Data storytelling: The essential data science skill everyone needs. <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>.
- [Escuela de Medicina de la Pontificia Universidad Católica de Chile, 2023] Escuela de Medicina de la Pontificia Universidad Católica de Chile (2023). Investigadores UC desarrollan estudio sobre reducción del estigma y salud mental desde la atención primaria de salud. <https://medicina.uc.cl/noticias/investigadores-uc-desarrollan-estudio-sobre-reduccion-del-estigma-y-salud-mental-desde-la-atencion-primaria-de-salud/>. Visitado el 8 de noviembre de 2023, 12:00.
- [Gomez-Adorno *et al.*, 2024] Gomez-Adorno, H., Bel-Enguix, G., Sierra, G., Barajas, J.-C., y Álvarez, W. (2024). Machine learning and deep learning sentiment analysis models: Case study on the sent-covid corpus of tweets in Mexican Spanish. En *Informatics*, volumen 11, p. 24. MDPI.
- [Hossain *et al.*, 2025] Hossain, M. M., Hossain, M. S., Mridha, M., Safran, M., y Alfarhood, S. (2025). Multi task opinion enhanced hybrid bert model for mental health analysis. *Scientific Reports*, 15(1):3332.

- [Instituto Nacional de la Juventud, 2019] Instituto Nacional de la Juventud (2019). Sondeo: Salud mental. principales resultados - jóvenes entre 15 y 29 años. https://www.injuv.gob.cl/sites/default/files/resultados_sondeo_salud_mental.pdf.
- [Ipsos, 2024] Ipsos (2024). La salud mental se considera como el principal problema de salud a nivel mundial y Chile lidera el ranking con un 69%. <https://www.ipsos.com/es-cl/la-salud-mental-se-considera-como-el-principal-problema-de-salud-nivel-mundial-y-chile-lidera-el>. Monitor de Servicios de Salud de Ipsos 2024, Noticias y Eventos.
- [Keim et al., 2013] Keim, D., Qu, H., y Ma, K.-L. (2013). Big-data visualization. *IEEE computer graphics and applications*, 33(4):20–21.
- [Khder, 2021] Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3).
- [Li et al., 2019] Li, J., Xu, Q., Shah, N., y Mackey, T. K. (2019). A machine learning approach for the detection and characterization of illicit drug dealers on instagram: Model evaluation study. *Journal of Medical Internet Research*, 21(6):e13803.
- [Ministerio de Salud, 2019] Ministerio de Salud (2019). Minsal lanza primera guía para la prevención del suicidio en adultos mayores y los insta a sumarse al #exigeelamarillo. <https://www.minsal.cl/minsal-lanza-primera-guia-para-la-prevencion-del-suicidio-en-adultos-mayores-y-los-insta-a-sumarse-al-exigeelamarillo/>. Visitado el 9 de noviembre de 2023, 15:00.
- [Ministerio de Salud, 2023] Ministerio de Salud (2023). Resumen ejecutivo: Primeros resultados. encuesta nacional de salud, sexualidad y género (enssex) 2022–2023. primera entrega de resultados. <https://epi.minsal.cl/resultados-encuestas/>. Primera entrega publicada en octubre de 2023.
- [Ministerio de Salud, 2025] Ministerio de Salud (2025). Informe encuesta nacional de salud, sexualidad y género 2022–2023. informe de resultados: Salud mental y género. <https://epi.minsal.cl/resultados-encuestas/>. Santiago de Chile, marzo 2025.
- [Moine et al., 2011] Moine, J. M., Gordillo, S. E., y Haedo, A. S. (2011). Análisis comparativo de metodologías para la gestión de proyectos de minería de datos. En *XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011)*.
- [Molina, 2021] Molina, D. (2021). Qué es el periodismo de datos y herramientas útiles. <https://www.iebschool.com/hub/big-data-y-el-periodismo-de-datos-big-data/>.
- [Observatorio de Salud Mental Chile, 2024] Observatorio de Salud Mental Chile (2024). Resultados octubre 2024 n°20.240: Salud mental y psicosocial. Informe técnico, Observatorio de Salud Mental Chile. Estudio Salud Mental en Chile.

- [Persson, 2019] Persson, E. (2019). Evaluating tools and techniques for web scraping.
- [Pérez et al., 2023] Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D. A., Luque, F., Alemany, L. A., y Martínez, M. V. (2023). pysentimiento: A python toolkit for opinion mining and social nlp tasks.
- [Sadiku et al., 2016] Sadiku, M., Shadare, A. E., Musa, S. M., Akujobi, C. M., y Perry, R. (2016). Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, 2(12):11-16.
- [Solis-Soto et al., 2019] Solis-Soto, M. T., Burrone, M. S., Reginatto, G., Sapag, J. C., y Alvarado, R. (2019). Stigma towards people with mental disorders: perceptions of devaluation and discrimination in a sample of chilean workers. *Salud mental*, 42(5):207-215.
- [Superintendencia de Salud, 2023] Superintendencia de Salud, D. d. E. y. D. (2023). Evaluación de la ley 21.331 de salud mental en Chile: Cobertura financiera y licencias médicas en isapres abiertas. https://www.superdesalud.gob.cl/documentacion/666/articulos-23775_recurso_1.pdf. Documento de trabajo publicado en agosto de 2023.
- [Superintendencia de Seguridad Social, 2021] Superintendencia de Seguridad Social (2021). Informe estadístico 2020 de licencias médicas en Chile. <https://www.suseso.cl/605/w3-article-652634.html>. Visitado el 8 de octubre de 2023, 16:30.
- [Superintendencia de Seguridad Social, 2023] Superintendencia de Seguridad Social (2023). Adelanto de cifras de informe de seguridad y salud en el trabajo. 67% de diagnósticos de enfermedades profesionales en Chile corresponden a salud mental. <https://www.suseso.cl/605/w3-article-705997.html>. Noticia institucional.
- [Superintendencia de Seguridad Social, 2025] Superintendencia de Seguridad Social (2025). Estadísticas de accidentabilidad 2024: Informe anual de seguridad y salud en el trabajo. <https://www.suseso.cl/607/w3-article-755198.html>. Informe estadístico sobre accidentes del trabajo y enfermedades profesionales, Ley N°16.744.
- [Vicente et al., 2004] Vicente, B., Kohn, R., Rioseco, P., Saldivia, S., Baker, C., y Torres, S. (2004). Population prevalence of psychiatric disorders in Chile: 6-month and 1-month rates. *The British journal of psychiatry : the journal of mental science*, 184:299-305.
- [World Health Organization, 2014] World Health Organization (2014). Mental health atlas 2017 country profile: Chile. <https://www.who.int/publications/m/item/mental-health-atlas-2014-country-profile-chile>. Mental Health, Brain Health and Substance Use (MSD), Technical document.
- [World Health Organization, 2017] World Health Organization (2017). Mental health atlas 2017 country profile: Chile. <https://www.who.int/publications/m/item/mental-health-atlas-2017-country-profile-chile>. Mental Health, Brain Health and Substance Use (MSD), Technical document.

[World Health Organization, 2022a] World Health Organization (2022a). Mental health atlas 2020 country profile: Chile. <https://www.who.int/publications/m/item/mental-health-atlas-chl-2020-country-profile>. Mental Health, Brain Health and Substance Use (MSD), Technical document.

[World Health Organization, 2022b] World Health Organization (2022b). Mental health: Strengthening our response. <https://www.who.int/es/news-room/fact-sheets/detail/mental-health-strengthening-our-response>. Visitado el 8 de octubre de 2023, 16:00.