UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

PERCEPCIÓN DE LA CALIDAD DEL SERVICIO EN EL SISTEMA DE SALUD MEDIANTE ANÁLISIS DE OPINIÓN EN BASES DE DATOS DISPONIBLES EN LA WEB

POR

Luis Sebastián Saavedra Acuña

Informe de Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción para optar al título profesional de Ingeniero Civil Biomédico

Profesor Guía
Rosa Liliana Figueroa Iturrieta

Comisión
Jean Paul Navarrete
Mario Medina

Junio 2025
Concepción (Chile)

Resumen

El presente trabajo tuvo como objetivo analizar la percepción de la calidad del servicio de salud en Chile a partir de opiniones de usuarios recopiladas en plataformas públicas en línea, utilizando técnicas de Procesamiento de Lenguaje Natural (PLN) y aprendizaje automático. La motivación principal radica en la falta de métodos automatizados de análisis de opinión en el contexto chileno, en contraste con la creciente disponibilidad de comentarios públicos en plataformas digitales. A través de web crawling automatizado, se recolectaron más de 2.000 opiniones de seis hospitales públicos y seis clínicas privadas, seleccionadas en base a criterios de tamaño poblacional y distribución geográfica.

Posteriormente, los textos fueron sometidos a un exhaustivo proceso de preprocesamiento, incluyendo la eliminación de acentos, normalización, tokenización y vectorización, para garantizar su correcta interpretación por los modelos analíticos. Para la clasificación de sentimientos, se implementaron tres algoritmos supervisados: Naive Bayes, Máquina de Vectores de Soporte (SVM) y una Red Neuronal Simple, utilizando la calificación numérica adjunta a cada opinión como criterio de etiquetado. Se decidió eliminar la clase neutra debido a su escasa representación, transformando el problema en una clasificación binaria entre sentimientos positivos y negativos. Los resultados mostraron que Naive Bayes obtuvo el mejor desempeño en hospitales, mientras que SVM lideró en clínicas, alineándose en parte con lo reportado en la literatura especializada.

En la segunda etapa, se aplicaron métodos de modelado de tópicos mediante aprendizaje no supervisado, utilizando Latent Dirichlet Allocation (LDA) y Factorización de Matrices No Negativas (FMN). Ambos enfoques permitieron identificar patrones temáticos recurrentes en las opiniones, tales como tiempos de espera, trato del personal y calidad de la atención, siendo FMN el que logró una mejor definición y segmentación de los tópicos, especialmente en los subconjuntos de opiniones negativas.

En conjunto, los resultados obtenidos evidencian que el uso de técnicas de PLN y machine learning puede automatizar de manera efectiva el análisis de la percepción ciudadana sobre el sistema de salud, proporcionando información valiosa para la mejora continua del servicio. Asimismo, el algoritmo desarrollado presenta potencial para ser implementado en sistemas de monitoreo en tiempo real, permitiendo a las instituciones de salud detectar tendencias en la opinión de los usuarios de manera temprana y oportuna, favoreciendo la toma de decisiones basada en evidencia.

Abstract

This work aimed to analyze the perception of healthcare service quality in Chile based on user opinions collected from public online platforms, utilizing Natural Language Processing (NLP) techniques and machine learning. The primary motivation lies in the lack of automated opinion analysis methods in the Chilean context, despite the increasing availability of public comments on digital platforms. Through automated web crawling, more than 2.000 opinions were collected from six public hospitals and six private clinics, selected based on population size and geographical distribution criteria.

The texts underwent an extensive preprocessing phase, including accent removal, normalization, tokenization, and vectorization, to ensure their correct interpretation by the analytical models. For sentiment classification, three supervised algorithms were implemented: Naive Bayes, Support Vector Machine (SVM), and a Simple Neural Network, using the numerical rating associated with each opinion as a labeling criterion. Due to its low representation, the neutral class was eliminated, transforming the problem into a binary classification between positive and negative sentiments. The results showed that Naive Bayes achieved the best performance in hospitals, while SVM outperformed the other models in clinics, partially aligning with findings reported in the literature.

In the second stage, topic modeling techniques based on unsupervised learning were applied, using Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). Both approaches enabled the identification of recurring thematic patterns in the opinions, such as waiting times, staff treatment, and quality of care, with NMF achieving a better definition and segmentation of topics, especially among negative opinions.

Overall, the results demonstrate that applying NLP and machine learning techniques can effectively automate the analysis of public perception of the healthcare system, providing valuable information for continuous service improvement. Furthermore, the developed algorithm shows potential for real-time monitoring applications, allowing healthcare institutions to detect emerging trends in user opinions in a timely manner and to support evidence-based decision-making.

Tabla de contenidos

Índice de Tablas

Índice de Figuras

Capítulo 1. Introducción

1.1 Introducción General

El actual sistema de salud vigente en Chile se caracteriza principalmente por estar dividido en dos. El sistema público, administrado por el Fondo Nacional de Salud (FONASA), y el sistema privado, administrado por las Instituciones de Salud Previsional (ISAPRE). Por su parte, FONASA brinda atención médica a la mayoría de la población chilena (alrededor de un 85 % de la población), mientras que ISAPRE se encuentra limitado a aquellos sectores de mayores ingresos capaces de costearlos (alrededor de un 15 % de la población). Esta dualidad en el sistema de salud ha generado una percepción de desigualdad en cuanto a calidad de la atención, tiempos de espera y acceso a servicios médicos especializados en ambos sectores, lo que ha abierto el debate a si el sistema de salud en su conjunto es capaz de brindar una buena atención [1].

A junio de 2024, se estima que la cantidad de personas en lista de espera corresponde a un total de 2.555.918 registros [2]. Esta lista de espera podría tener sus causas en el déficit de médicos especialistas que experimenta el sector de salud y su distribución territorial. La superintendencia de salud estima que alrededor del 59.7 % de médicos especialistas se encuentra en la región Metropolitana [3]. Lo anterior se suma a la falta de recursos; de acuerdo a los datos de la Organización para la Cooperación y el Desarrollo Económico (OECD), se estima que alrededor del 9 % del PIB nacional es destinado a salud, siendo inferior al de otros países de la región, como Uruguay o Cuba [4]. Los problemas mencionados, más la falta de atención, han generado que en la población se forme una opinión negativa sobre la atención de salud en Chile, sobre todo si hablamos del sector público [1].

En contraste, otras estadísticas relativas a salud dadas por la OECD indican que Chile se ubica como líder en la región en cuanto a salud de la población, ya que es el país de Latinoamérica con la mayor esperanza de vida, menor mortalidad infantil y mayor tasa de vacunación, estando además muy cercano al promedio de los países de la OECD [4]. Estos datos valoran de forma positiva el sistema de salud chileno, en comparación con el de otros países de la región.

Luego, surge la pregunta ¿es el sistema de salud chileno bueno o no? ¿cómo lo perciben sus usuarios? Para responder estas preguntas, se pueden dar argumentos tanto a favor como en contra, pero hay un hecho innegable: quien tiene la última palabra al respecto son los pacientes que se atienden en el sistema de salud. Dicho esto, las opiniones de los pacientes se han vuelto una fuente clave para evaluar el desempeño

de ambos sectores; es por ello que muchos hospitales realizan encuestas internas, sin embargo, estas son de carácter voluntario y pueden estar ligadas a sesgo del propio hospital [5], es por ello que muchos usuarios han preferido escribir su opinión en la web.

1.2 Problemática

En sitios web como Google Maps y Doctoralia, así como en redes sociales como Facebook y X, es común encontrar una amplia variedad de opiniones en texto libre sobre la atención en salud. Estos comentarios contienen información valiosa que puede brindar retroalimentación, tanto sobre los aspectos donde el sistema de salud está funcionando bien, como sobre aquellos que generan descontento o donde la atención podría mejorar. Aprovechar estos datos representa una oportunidad para identificar áreas de mejora y potenciar la calidad de los servicios de salud.

Si bien es posible analizar estos datos de forma manual, realizar esta tarea resulta inviable debido a la gran cantidad de datos disponibles y la subjetividad inherente a los comentarios, lo que lo convierte en un proceso complejo y que demanda mucho tiempo. Para dar un ejemplo, el Ministerio de Hacienda cada año realiza encuestas donde principalmente recopila datos numéricos y algunas opiniones de texto libre mediante entrevistas a los usuarios; sin embargo, estas opiniones se analizan de forma manual, y debido a la complejidad de este método. Lo anterior se suma a la cobertura que alcanzan estos métodos, que algunas veces no son suficientes para obtener conclusiones representativas. Un ejemplo de esto es la versión 2022 de la encuesta del Ministerio de Hacienda, que entrevistó solo a 6 personas [6].

Es aquí donde el Procesamiento de Lenguaje Natural (PLN) surge como una herramienta esencial. El PLN permite automatizar el análisis de grandes volúmenes de texto, extrayendo patrones y sentimientos que pueden ofrecer una visión más clara y estructurada sobre la percepción de los usuarios del sistema de salud [7][8][9][10]. Adicionalmente, si las bases de datos utilizadas en los análisis son amplias, se puede lograr mayor cobertura.

El PLN se utiliza igualmente para realizar análisis de sentimiento o polaridad de texto libre. Estos análisis han demostrado que es posible identificar tendencias en las opiniones, clasificando los comentarios en términos de satisfacción o insatisfacción [11]. De esta forma, se puede detectar de manera eficiente en qué áreas del sistema de salud existen problemas recurrentes, así como dónde se están generando valor y buenas experiencias. Dado lo anterior, este proyecto de memoria de título propone utilizar técnicas de análisis de polaridad de textos libres, junto con datos extraídos de la web para responder la pregunta sobre ¿qué opinan los usuarios sobre el sistema de salud chileno?.

1.3 Objetivos

1.3.1 Objetivo general

Desarrollar un algoritmo basado en técnicas de PLN en conjunto con técnicas de aprendizaje automático que permita realizar la clasificación de tópicos y un análisis de polaridad de las opiniones de los usuarios del sistema de salud chileno, disponibles en la web.

1.3.2 Objetivos específicos

- Realizar un análisis de la literatura existente en análisis de sentimientos y opiniones con datos de la web.
- Definir la metodología a utilizar para extraer el set de datos y realizar una clasificación de tópicos.
- Aplicar análisis de sentimientos para conocer la opinión de los usuarios del sistema de salud Chileno.

1.4 Alcances y limitaciones

Dados los recursos computacionales y bibliográficos con los que se cuenta, los recursos y limitaciones son los siguientes:

- El equipo que se utilizará es un Acer Nitro 5 an515-54 con procesador Intel® Core™ i5 de 9° generación y memoria RAM de 32 GB.
- Las opiniones a utilizar son públicas y se encuentran disponibles a través del api de google maps. Para proteger la privacidad de las personas e instituciones no serán utilizados los nombres de personas o instituciones.
- Las técnicas de PLN a utilizar se escogerán en base a los resultados obtenidos en la bibliografía revisada.
- El software que se utilizará es Spyder versión 5.4.3, en el lenguaje Python 3.11.4.
- La base de datos fue consultada el 18/11/2024.

1.5 Metodología de Trabajo

1. Se realizará un webcrawling con el fin de recopilar la base de datos disponibles en la web.
2. Mediante la calificación de cada opinión, se etiquetarán automaticamente los datos según el sentimiento de la calificación.
3. Se entrenarán algoritmos de aprendizaje supervisado sobre la base de datos etiquetada y se obtendran métricas relevantes.
4. Se realizará modelado de tópicos mediante algoritmos de aprendizaje no supervisado y se evaluará la calidad de los temas generados.
5. Se analizarán los resultados y se obtendran conclusiones al respecto.

1.6 Temario

1. Capitulo 1: Contextualiza la problemática del sistema de salud chileno, plantea los objetivos del estudio y describe el alcance de la investigación.
2. Capitulo 2: Revisa la literatura sobre encuestas de satisfacción del paciente y técnicas de PLN, destacando métodos de análisis de sentimiento y modelado de tópicos.
3. Capitulo 3: Expone los conceptos clave del sistema de salud, el PLN y las técnicas de clasificación de texto supervisadas y no supervisadas.
4. Capitulo 4: Describe las técnicas y herramientas utilizadas para el webcrawling, el preprocesamiento de datos y el análisis de opiniones.
5. Capitulo 5: Presenta y describe los resultados obtenidos mediante matrices, graficos, y nubes de palabras.
6. Capitulo 6: Se discuten los resultados obtenidos, se concluye como estos cumplieron los objetivos del proyecto y se plantea el trabajo futuro a realizar.

Capítulo 2. Estado del Arte

En este capítulo se expone el estado actual en el que se encuentran las encuestas de satisfacción al cliente en el mundo, con especial énfasis en Chile, y el estado actual de las distintas técnicas de PLN utilizadas en análisis de opinión.

2.1 Encuestas de Satisfacción del paciente

La opinión de los pacientes es importante para los prestadores de salud, pues, así como en todos los servicios, una buena calificación de parte de los clientes es un indicador de prestigio, además de ser de vital importancia cuando se busca mejorar la atención en salud. Usualmente los usuarios suelen buscar los centros de salud mejor evaluados y recomendar a otras personas, dejando su opinión en sitios web públicos o en redes sociales. Ahora bien, ¿Cómo se puede medir la satisfacción de un paciente? Para esto se han creado varias encuestas, como lo son la encuesta Press Ganey, una de las primeras creadas para conocer la satisfacción del paciente [7], la encuesta Friend and Family Test (FFT), empleada por el Servicio Nacional de Salud de Inglaterra [12], la encuesta Patient-Reported Experience Measures (PREM), una encuesta desarrollada en Países Bajos la cual es realizada directamente a los pacientes [13], la encuesta Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), obligatoria en todos los hospitales de Estados Unidos [9], la encuesta Service Quality (SERVQUAL), una encuesta que mide la satisfacción de los usuarios con un determinado servicio la cual esta dividida en 5 dimensiones [14], etc. También se puede presentar el caso de que el hospital en cuestión diseñe su propia encuesta personalizada basada en sus servicios, aunque estas suelen tener cierto sesgo asociado del propio hospital [5].

Para el caso de Chile, desde el año 2015, el Ministerio de Hacienda realiza la encuesta de Medición de Satisfacción Usuaria (MESU), dicha encuesta tiene por objetivo medir qué tan satisfechos están los usuarios con distintas instituciones estatales, y mediante esto, identificar oportunidades de mejora. Para el caso particular de las instituciones de salud, la encuesta se realiza mediante llamadas telefónicas, las preguntas realizadas son principalmente preguntas con respuestas preestablecidas, o preguntas cuantitativas, donde se le pregunta al paciente sobre un servicio prestado, y se le solicita que le dé una calificación del 1 al 7. Los resultados de estas encuestas se obtienen mediante análisis cuantitativos y cualitativos a las preguntas abiertas de éstas. El análisis de estas preguntas se realiza de forma manual, lo que requiere de conocimiento y tiempo. Por lo mismo, no es de extrañar que en la encuesta realizada el año 2022, solo

hayan participado 6 personas [6].

La escasa cantidad de información en forma de texto libre proporcionada por encuestas oficiales da pie a la necesidad de buscar otras fuentes de donde extraer datos; afortunadamente, en sitios web y en redes sociales existe una inmensa cantidad de información asociada a las experiencias de usuarios. Diferentes estudios demostraron que el uso de este tipo de información es bastante conveniente para el análisis debido a la gran cantidad de información que se puede extraer de estos sitios, la cual abarca todo tipo de tópicos y es proveniente de personas de todas las edades, con condiciones de salud de lo más diversas [15][9][5]. Otros estudios coinciden en tomar tanto los datos provenientes de encuestas oficiales como los provenientes de sitios web y redes sociales, con el fin de formar una "nube" o "tormenta" de información del paciente, que permita obtener la mayor cantidad de información posible [5][10], y de esta forma obtener una visión más profunda y amplia sobre la satisfacción de los usuarios en el sistema de salud. Esta visión permitiría no solo identificar problemas recurrentes, sino también proponer mejoras más efectivas, basadas en la experiencia de los pacientes.

2.2 Tecnicas de Procesamiento de Lenguaje Natural aplicadas a análisis de opinión de usuarios

Las técnicas de PLN permiten el análisis de grandes cantidades de datos no estructurados, facilitando la extracción de información que no sería accesible a través de datos cuantificables [12]. Mediante el uso de técnicas como el análisis de sentimientos, es posible identificar si un comentario tiene un sentimiento positivo, negativo, mixto o neutral. Para realizar este análisis, lo más típico es utilizar métodos basados en diccionarios de palabras. Estos métodos asignan un valor de polaridad a las palabras en función de si expresan un sentimiento positivo, negativo o neutral. Herramientas como Vader o SentiWordNet son ejemplos de estos enfoques, y se utilizan frecuentemente para analizar sentimientos en textos como comentarios de redes sociales o reseñas de servicios de salud [14].

Además, al combinar esta técnica con la clasificación de tópicos, que consiste en agrupar automáticamente los textos en temas relacionados, se puede identificar en qué temas específicos hay una mayor concentración de sentimientos positivos o negativos. Esto permite analizar de manera más eficiente qué aspectos generan más satisfacción o insatisfacción en los usuarios. Un estudio desarrollado por Nawab et al. mostró que la mayoría de los comentarios positivos en un hospital estaban asociados a la relación paciente-médico y paciente-enfermera, demostrando la importancia de estos profesionales. Asimismo, se evidenció que la mayoría de los comentarios negativos estaban relacionados con la temperatura de las habitaciones, un resultado curioso, dado que este no era un tema muy frecuente [7].

Cabe destacar que para el proceso de clasificación de tópicos se pueden utilizar algoritmos basados en aprendizaje supervisado, como Naive Bayes (NB) o Máquina de Vectores de Soporte (SVM), los cuales requieren ser entrenados con datos etiquetados; en aprendizaje no supervisado, que trabaja con datos no etiquetados, como Latent Dirichlet Allocation (LDA) o Factorización de Matrices No Negativas (FMN); y en aprendizaje semi-supervisado, que combina un algoritmo no supervisado con algunos datos etiquetados [12]. Sin embargo, si se desea utilizar un enfoque supervisado, es crucial contar con una base de datos correctamente etiquetada, ya que el aprendizaje supervisado puede replicar sesgos presentes en los datos de entrenamiento, lo que afectaría la precisión y equidad de la clasificación de tópicos [11][13][9].

Con una base de datos bien etiquetada, los modelos de aprendizaje supervisado, como NB o SVM, han demostrado ser bastante efectivos en comparación con la clasificación manual. Por ejemplo, Buchem et al. utilizaron Naive Bayes para realizar clasificación automática de tópicos, y también realizaron una clasificación manual de referencia; la coincidencia entre ambas superó el 90 % [13]. Asimismo, Ötles et al. probaron distintos métodos de clasificación, incluyendo Naive Bayes, regresión logística y bosques aleatorios, y concluyeron que SVM era el modelo que mostraba los mejores resultados, con una precisión del 83 % para clasificaciones donde se tenían pocos temas [16].

Existen otros algoritmos de clasificación más sofisticados como Redes Neuronales Artificiales (RNA), Redes Neuronales Convergentes (RNC), o Redes Neuronales Recurrentes (RNR), pero estos en comparación con algoritmos tradicionales, no han mostrado mejores resultados cuando se trabaja con bases de datos no muy extensas. Harrison & Sidey-Gibbons realizaron clasificación usando regresión logística, SVM, y RNA. Los resultados mostraron que SVM fue el más preciso con 72 % de precisión, concluyendo que RNA no ofrecía una ventaja significativa en comparación con algoritmos tradicionales, y que para mejorar su precisión sería necesario trabajar con una base de datos mucho más grande y realizar ajustes refinados en el sistema de la red [17]. Por otro lado, cuando se dispone de una gran base de datos y se realizan los ajustes correspondientes, los modelos basados en redes neuronales han mostrado mejores rendimientos que los modelos tradicionales. Shah et al utilizaron un modelo de RNC, además de métodos tradicionales como NB y SVM para el análisis de opiniones. El modelo de RNC logró una precisión de 97.75 %, superando considerablemente a los modelos de NB y SVM que rondaron entre el 84 % y 89 %, respectivamente. Sin embargo, este rendimiento óptimo de las redes neuronales solo fue posible gracias al uso de un amplio conjunto de datos de 55,738 reseñas, resaltando la necesidad de grandes volúmenes de datos para aprovechar todo su potencial [18].

De manera similar, un estudio realizado por Lavanya & Sasikala destaca la efectividad de las redes neuronales con arquitecturas híbridas, como RNC-RNR, para la clasificación de texto en el ámbito de la

salud. Estas arquitecturas son altamente eficaces para capturar tanto patrones locales como dependencias contextuales a largo plazo, logrando un rendimiento superior al de métodos tradicionales. Estos modelos de aprendizaje profundo no solo alcanzan mayores niveles de precisión, sino que también son más eficientes en la extracción de patrones semánticos complejos y en la detección de emociones en redes sociales de salud. No obstante, se enfatiza que el rendimiento óptimo de estos modelos depende del uso de grandes conjuntos de datos y un ajuste cuidadoso de los hiperparámetros, lo que refuerza la necesidad de bases de datos extensas y equilibradas para maximizar su efectividad [19].

Sin embargo, una de las principales limitaciones de la clasificación supervisada es que se realiza sobre un conjunto de tópicos predefinidos, por lo que el algoritmo no puede descubrir nuevos temas. Solo clasifica los textos en los tópicos definidos previamente por el usuario. Por el contrario, los algoritmos de aprendizaje no supervisado, como LDA, no requieren de tópicos predefinidos y son capaces de modelar y descubrir tantos tópicos como el algoritmo identifique en los datos. Este enfoque es especialmente útil en el ámbito de la salud, donde se busca comprender temas emergentes en grandes volúmenes de datos no etiquetados. Harrison & Sidey-Gibbons demostraron que LDA es efectivo para identificar temas latentes en los comentarios de pacientes, permitiendo a los investigadores descubrir patrones y problemas recurrentes sin intervención manual. De este artículo se desprende que el uso de LDA es útil en estudios exploratorios, donde se desea entender los aspectos más mencionados por los pacientes en sus opiniones y experiencias clínicas [17]. Por otro lado, Cammel et al. utilizaron FMN, que descompone una matriz grande en dos matrices más pequeñas cuyas entradas son todas no negativas. Este método permite identificar patrones ocultos en los datos, como temas o tópicos en textos. El uso de esta técnica para el modelado de tópicos resultó en un total de 127 tópicos. [11].

En algunos casos, el proceso de clasificación de temas puede requerir contexto adicional para comprender por qué ciertos tópicos presentan una mayor concentración de comentarios negativos. Este problema es menos común en los algoritmos no supervisados, ya que la cantidad de tópicos descubiertos es mayor y tienden a ofrecer más detalles. Sin embargo, en los algoritmos supervisados, que trabajan con un conjunto limitado de temas predefinidos, puede ser necesario utilizar n-gramas para proporcionar más contexto. En un estudio de Vashishtha & Kapoor, se identificó que muchos comentarios negativos estaban asociados con el tema "página web". El análisis de n-gramas permitió descubrir que el problema principal era la dificultad de los usuarios para ingresar a dicha página web [8].

En los últimos años, las técnicas de procesamiento de lenguaje natural han avanzado significativamente, permitiendo la clasificación automática de grandes volúmenes de datos textuales y la extracción de patrones de opiniones que, de otro modo, pasarían desapercibidos. Sin embargo, como han mostrado varios estudios, aún persisten ciertos desafíos. Uno de ellos es la dificultad para garantizar la precisión

de los modelos en diferentes contextos culturales y lingüísticos, como lo señala Shaw et al. al analizar comentarios en redes sociales, donde las variaciones idiomáticas y contextuales pueden influir en los resultados [15]. Además, aunque los enfoques no supervisados, como el modelado de temas, han demostrado ser efectivos en la clasificación de tópicos, existe el riesgo de generar demasiados temas que pueden complicar la interpretación de los resultados [11]. A pesar de estos desafíos, las oportunidades que ofrece el PLN en el campo de la salud son vastas, desde la mejora de la experiencia del paciente hasta la optimización de la atención médica, como lo evidencian las investigaciones de Nawab et al. y Buchem et al [7][13]. En el futuro, la combinación de técnicas como el análisis de sentimientos, la clasificación de tópicos y el uso de n-gramas seguirá siendo fundamental para extraer información significativa de datos no estructurados, ayudando a comprender mejor las necesidades y preocupaciones de los pacientes en contextos tan diversos como lo podría ser el sistema de salud chileno.

2.3 Discusión

En función de la revisión bibliográfica realizada, se decidió aplicar una combinación de métodos de aprendizaje supervisado y no supervisado para el análisis de las opiniones de los pacientes en este estudio. Esta decisión responde tanto a las características de la base de datos disponible como a los objetivos específicos de análisis de sentimiento y clasificación de tópicos.

La base de datos cuenta con 2.041 opiniones, de las cuales el sentimiento ha sido etiquetado indirectamente, ya que cada opinión incluye una clasificación del 1 al 5. Este sistema de calificación permite clasificar las opiniones como positivas (4 y 5), neutras (3) o negativas (1 y 2), lo cual hace viable el uso de un enfoque de aprendizaje supervisado para el análisis de sentimiento. Este tipo de análisis supervisado tiene la ventaja de ofrecer mayor precisión y control sobre las categorías de sentimiento al estar entrenado con datos previamente etiquetados. Según la literatura, métodos supervisados como Naive Bayes y SVM han demostrado buenos resultados en este tipo de tareas, especialmente cuando el objetivo es clasificar en categorías definidas. La literatura no recomienda el uso de Redes Neuronales debido a que estas solo funcionan de forma adecuada cuando se tiene una base de datos de gran tamaño; no obstante, de todos modos se empleará un modelo de red neuronal simple para comprobar si los resultados obtenidos son acordes con los resultados de las investigaciones realizadas previamente.

Para la clasificación de tópicos, la base de datos no cuenta con etiquetas predefinidas. Dado esto, se seleccionó un enfoque de aprendizaje no supervisado que permita identificar temas latentes sin necesidad de etiquetas previas. En este contexto, se emplearán dos métodos complementarios: LDA y FMN. LDA fue seleccionado debido a su eficacia comprobada para descubrir temas ocultos en datos textuales exten-

sos y no etiquetados, lo cual es esencial para comprender las preocupaciones emergentes de los pacientes en diversas áreas de la atención médica. Este modelo ha sido ampliamente validado en la literatura, especialmente en el ámbito de la salud, por su capacidad para ofrecer una visión exploratoria de los temas subyacentes en grandes volúmenes de datos no etiquetados.

No obstante, también se incorporará FMN debido a su capacidad para descomponer la matriz de documentos en factores interpretables, aplicando restricciones de no negatividad, lo que facilita la obtención de tópicos más claros y compactos. Estudios previos han demostrado que FMN puede complementar a LDA al proporcionar una representación más directa y menos difusa de los temas emergentes, mejorando así la interpretabilidad de los resultados.

Capítulo 3. Marco Teórico

En esta sección, se presentan los conceptos más relevantes que se van a tratar a lo largo de la investigación, relativos a las instituciones de salud, las encuestas de satisfacción y las distintas técnicas de PLN utilizadas para su análisis:

3.1 Sistema de Salud

El sistema de salud es el conjunto de organizaciones, instituciones y recursos cuyo objetivo principal es mejorar la salud de la población. El sistema de salud chileno es mixto, combinando elementos del sector público y privado, y ha evolucionado a lo largo de su historia para responder a las necesidades de la población y los cambios epidemiológicos. En Chile, el sistema actual es el resultado de diversas reformas que integran tres modelos principales: el modelo de Seguridad Social, el modelo de Servicio Nacional de Salud y el modelo Privado. El sistema de seguridad social se financia mediante contribuciones obligatorias y control estatal, mientras que el modelo de servicio nacional es financiado a través de impuestos y gestionado principalmente por entidades públicas. Por otro lado, el modelo privado se basa en la contribución individual a aseguradoras privadas, lo que excluye el concepto de solidaridad presente en los otros dos sistemas [20].

3.2 Procesamiento de Lenguaje Natural (PLN)

Es una rama de la inteligencia artificial que se encarga de la interacción entre computadoras y el lenguaje humano. Su objetivo es hacer que las máquinas entiendan, interpreten y generen el lenguaje natural de manera que sea útil para diversas aplicaciones, como la clasificación de textos, la traducción automática, la minería de opiniones, entre otros. Para que una computadora pueda comprender el lenguaje humano, es necesario transformar los textos en una representación que pueda ser procesada [21]. A continuación, se describen algunos de los pasos fundamentales que permiten esta transformación:

- Normalización de texto: Implica procesos como convertir todo el texto a minúsculas, eliminar caracteres especiales o acentos, para que las palabras puedan ser comparadas de manera uniforme.
- Eliminación de Stopwords: Las stopwords son palabras comunes como ’el’, ’la’, ’de’, ’en’, que

no aportan mucho valor semántico en el análisis del texto. Estas palabras suelen eliminarse para reducir el ruido en el procesamiento.

- Tokenización: Consiste en dividir el texto en unidades más pequeñas, comúnmente llamadas tokens, que pueden ser palabras, oraciones o párrafos, dependiendo del nivel de granularidad que se desee.
- Stemming: El stemming reduce las palabras a su raíz o forma base. Por ejemplo, las palabras 'corriendo' y 'corrió' pueden reducirse a 'corr'. Esto ayuda a agrupar palabras derivadas de una misma raíz para simplificar el análisis.
- Vectorización: Una vez preprocesado el texto, es necesario codificar las palabras como números. Esto puede hacerse mediante técnicas como Bag of Words o Term Frequency-Inverse Document Frequency (TF-IDF). Ambas técnicas transforman las palabras en una representación de espacio vectorial.

3.3 Análisis de Sentimiento

Técnica de PLN utilizada para identificar y extraer la emocionalidad de un texto. Funciona al clasificar palabras o frases en categorías como 'positivas', 'negativas' o 'neutrales'. Existen varios enfoques para realizar esta tarea, desde reglas basadas en diccionarios de palabras hasta modelos más avanzados de aprendizaje profundo [21]. Usualmente, se comienza con la tokenización y normalización del texto. Luego, un modelo basado en diccionarios de palabras, como Vader o SentiWordNet, analiza el texto para detectar palabras o frases asociadas con emociones. Para mejorar la precisión, se pueden utilizar técnicas como TF-IDF o Word2Vec para representar las palabras. Este método se utiliza para monitorear opiniones en redes sociales, analizar encuestas de satisfacción o evaluar comentarios de productos y servicios. En el contexto de la atención médica, permite conocer la percepción de los pacientes sobre aspectos como la calidad del servicio o la atención recibida [12][8][10].

3.4 Clasificación de texto

Es una técnica en la que un modelo asigna una o más categorías a un texto determinado. Es fundamental para tareas como el filtrado de spam, la clasificación de artículos de noticias y el análisis de comentarios de clientes. Los textos son preprocesados (normalización, tokenización, eliminación de stopwords) y luego representados en una forma adecuada para el algoritmo (como Bag of Words o TF-IDF).

En el caso del aprendizaje supervisado, como un clasificador NB o un SVM, el modelo es entrenado con ejemplos de textos etiquetados en diversas categorías [21]. El modelo aprende a reconocer patrones de palabras y características específicas de cada categoría, permitiendo así clasificar nuevos textos. En el ámbito hospitalario, la clasificación de texto puede ayudar a identificar los temas en los que los pacientes expresan más preocupación o satisfacción, como los tiempos de espera, la atención médica, o las condiciones del hospital. Esto ayuda a priorizar mejoras en áreas críticas basadas en los comentarios de los pacientes [8][11].

Por otro lado, en el aprendizaje no supervisado, el modelo no requiere ejemplos etiquetados. Los algoritmos no supervisados, como LDA o FMN, agrupan automáticamente los textos en temas relacionados basándose en patrones ocultos en los datos. Estos modelos son especialmente útiles cuando no se cuenta con datos etiquetados y se busca descubrir los temas latentes que surgen de manera natural en los textos. Por ejemplo, en un hospital, al analizar los comentarios de los pacientes con un enfoque no supervisado, se pueden descubrir nuevos temas que no habían sido previamente identificados, como preocupaciones inesperadas sobre la comida, la limpieza o la privacidad [11]. Estos modelos identifican los tópicos subyacentes y los asocian con los textos, permitiendo obtener una visión más amplia y menos sesgada de los temas que preocupan a los pacientes.

3.5 Modelos de Clasificación Supervisada

- **Naive Bayes (NB):** está basado en el teorema de Bayes, que se define de la siguiente manera:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \tag{3.5.1}$$

  Donde:

  - $P(C|X)$ es la probabilidad posterior de que la clase sea C dado el conjunto de características X.
  - $P(X|C)$ es la probabilidad condicional de las características X dado que la clase es C.
  - $P(C)$ es la probabilidad a priori de la clase
  - $P(X)$ es la probabilidad total de observar el conjunto de características X.

  Este algoritmo supone que las características (las palabras en este caso) son independientes entre sí, lo cual es una simplificación (de ahí el término 'Naive' o ingenuo). lo cual simplifica el cálculo $P(X|C)$ como el producto de las probabilidades de cada característica individual de la siguiente

forma:

$$P(X|C) = P(x_1|C) \cdot P(x_2|C) \cdot \cdots \cdot P(x_n|C)$$

A pesar de esta suposición simplificadora, el algoritmo puede funcionar bastante bien en muchos problemas de clasificación de texto [21].

- **Máquina de vectores de soporte (SVM):** funciona encontrando un 'hiperplano' que separa los datos en distintas clases con el mayor margen posible. La función objetivo de SVM es la siguiente:

$$\min \frac{1}{2}||w||^2 \tag{3.5.2}$$

Sujeto a la restricción de clasificación correcta para cada punto de datos:

$$y_i(w \cdot x_i + b) \geq 1, \quad \forall i$$

Donde:

- w es el vector normal al hiperplano.
- x_i son los puntos de datos.
- y_i son las etiquetas de clase (+1 o -1).
- b es el sesgo.

En el caso de la clasificación de texto, se representa cada documento como un punto en un espacio multidimensional, donde cada dimensión es una palabra o característica. El objetivo de SVM es encontrar una frontera entre diferentes clases (por ejemplo, sentimientos positivos y negativos) que maximice la distancia entre los ejemplos de cada clase [21].

- **Redes Neuronales Artificiales:** son modelos computacionales inspirados en la estructura del cerebro humano, compuestos por unidades llamadas neuronas artificiales que se organizan en capas. Cada neurona recibe entradas, realiza una operación matemática sobre ellas y transmite una salida hacia las siguientes capas. Estos modelos han sido ampliamente utilizados en tareas de PLN, como la clasificación de texto, el análisis de sentimientos, la detección de tópicos y la traducción automática.

  Una red neuronal feedforward básica está compuesta por tres tipos de capas: una capa de entrada, una o más capas ocultas, y una capa de salida. Cada conexión entre neuronas está asociada a un peso que se ajusta durante el entrenamiento para minimizar el error entre la salida esperada y la obtenida. Este ajuste se realiza mediante algoritmos de optimización como el descenso por gradiente y el método de retropropagación.

El funcionamiento básico de una neurona puede representarse con la siguiente ecuación:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{3.5.3}$$

Donde:

- x_i representa las entradas.
- w_i son los pesos asociados.
- b es el sesgo.
- f es la función de activación (por ejemplo, sen, sigmoide o tanh).
- y es la salida de la neurona.

El aprendizaje en una red neuronal consiste en encontrar los pesos w_i que minimicen una función de pérdida, como la entropía cruzada en problemas de clasificación [17][18]. Este proceso se realiza de forma iterativa utilizando un conjunto de datos etiquetados.

3.6 Modelos de Clasificación No Supervisada

- **Latent Dirichlet Allocation (LDA):** Es uno de los algoritmos más utilizados para el modelado de tópicos. LDA supone que:

  1. Para cada documento d, se genera una distribución de tópicos θ_d desde una distribución Dirichlet con parámetro α:

     $$\theta_d \sim \text{Dirichlet}(\alpha)$$

  2. Para cada tópico k, se genera una distribución de palabras ϕ_k desde una Dirichlet con parámetro η:

     $$\phi_k \sim \text{Dirichlet}(\eta)$$

  3. Para cada palabra $w_{d,n}$ en el documento d:

     a) Se elige un tópico $z_{d,n}$ desde la distribución de tópicos del documento:

     $$z_{d,n} \sim \text{Multinomial}(\theta_d)$$

     b) Se elige una palabra $w_{d,n}$ desde la distribución de palabras asociada al tópico $z_{d,n}$:

     $$w_{d,n} \sim \text{Multinomial}(\phi_{z_{d,n}})$$

El algoritmo analiza los datos textuales para identificar qué temas subyacentes (o tópicos) están presentes en los documentos. No requiere etiquetas previas, lo que lo hace ideal para descubrir temas ocultos en grandes colecciones de texto. El proceso de LDA consiste en asignar de manera iterativa palabras a tópicos y ajusta las asignaciones para maximizar la probabilidad de que el texto sea generado por una combinación de temas [22]. Esto permite identificar automáticamente los temas más relevantes, como los problemas que más mencionan los pacientes.

- **Factorización de Matrices No Negativas (FMN):** Este algoritmo descompone una matriz de datos $V \in \mathbb{R}^{m \times n}$ (por ejemplo, una matriz de términos por documentos) en dos matrices no negativas W y H, tal que:

$$V_{m \times n} \approx W_{m \times k} \cdot H_{k \times n} \tag{3.6.1}$$

Donde:

- V: matriz original de frecuencia de términos por documentos.
- W: matriz de términos por tópicos.
- H: matriz de tópicos por documentos.
- k: número de tópicos latentes, cumpliéndose que $k \ll m, n$.

En el caso del análisis de texto, FMN agrupa palabras y documentos en conjuntos que representan tópicos latentes, donde cada palabra contribuye de manera positiva a un tema. Esta técnica es útil para identificar patrones ocultos en los datos textuales, como comentarios de pacientes, y es especialmente útil cuando se busca una interpretación sencilla de los resultados. FMN se utiliza en la clasificación no supervisada porque permite descubrir relaciones entre palabras y documentos sin necesidad de etiquetas, proporcionando una representación clara de los tópicos subyacentes [22].

3.7 Análisis de n-gramas

Son secuencias de n elementos consecutivos de un texto. Un n-grama puede consistir en una secuencia de palabras o carácteres, dependiendo de la granularidad. Esta técnica es útil para analizar el contexto y las relaciones entre palabras en un texto. Primero, el texto se tokeniza, dividiéndolo en palabras o caracteres, luego, los n-gramas se generan tomando secuencias de tokens de longitud n (bigramas para 2 palabras, trigramas para 3 palabras, etc.). Esto permite capturar combinaciones de palabras que podrían tener un significado particular cuando se leen juntas, y que no son evidentes al analizar palabras aisladas [21].

En el análisis de comentarios de pacientes, los n-gramas permiten detectar combinaciones frecuentes de palabras que proporcionan más contexto sobre ciertos tópicos [8]. A continuación en la Fig.3.1, se muestra un ejemplo de n-gramas para la oración 'esto es un ejemplo' con distintos valores de n:

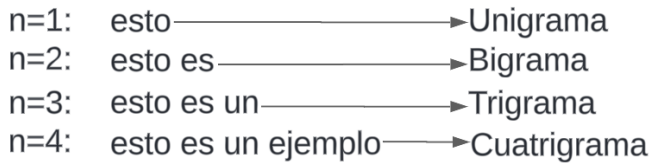


Figura 3.1: Ejemplo de n-gramas

3.8 Webcrawling

Técnica automatizada para recopilar grandes volúmenes de datos de la web. Se utiliza principalmente para extraer información relevante de múltiples páginas web de manera rápida y eficiente. Los web crawlers, o spiders, son programas que exploran automáticamente los sitios web siguiendo enlaces de una página a otra, mientras extraen datos como texto, imágenes, o enlaces. Una vez que el crawler descarga los datos de las páginas seleccionadas, se pueden procesar y analizar utilizando técnicas de PLN, para convertir la información no estructurada en datos útiles. En el contexto de análisis de comentarios de pacientes, el web crawling puede usarse para recopilar opiniones de diversas plataformas online, como foros, redes sociales o sitios de reseñas, y luego, procesar estos comentarios para obtener perspectivas sobre la experiencia de los pacientes [23]. A continuación, se mencionan algunos de los web crawlers más populares:

- Scrapy: permite definir spiders que exploran sitios web siguiendo enlaces y extrayendo información relevante según las reglas especificadas por el usuario. Utiliza técnicas como la extracción de datos estructurados (HTML, JSON) para obtener el contenido de interés, que luego se puede procesar usando técnicas de PLN [24].
- Selenium: herramienta de automatización que puede controlar navegadores web de manera programática, permite extraer datos de sitios dinámicos (como redes sociales o sitios con contenido cargado por JavaScript), automatizar pruebas web, o interactuar con formularios web para realizar acciones repetitivas [24].
- Chromedriver: API que actúa como intermediario entre el navegador Google Chrome y herramientas de automatización como Selenium. Permite controlar el navegador de forma automática

mediante scripts, simulando acciones humanas como hacer clic, desplazarse por páginas web o extraer información. Es desarrollado y mantenido por Chrome, y está diseñado específicamente para ejecutar comandos en navegadores Chrome de forma precisa y eficiente [24].

Capítulo 4. Metodología

En este capítulo se describen los distintos métodos y técnicas que se emplearon para desarrollar el algoritmo de análisis de opiniones, así como su implementación en el contexto de este estudio. Se detalla cada proceso, desde la obtención y preprocesamiento de los datos, hasta la implementación y ajuste de los modelos que se utilizaron para el análisis de sentimiento y la clasificación de tópicos.

4.1 Obtención de la Base de Datos

Para la obtención de la base de datos se empleó la biblioteca de Python Selenium, junto con Chromedriver como WebDriver, lo cual permitió automatizar y controlar el navegador Google Chrome de manera precisa para realizar el webcrawling. Selenium fue seleccionado debido a su flexibilidad y capacidad para interactuar con elementos dinámicos de las páginas web, tales como botones y ventanas emergentes, algo esencial en este caso, ya que los datos estaban ubicados en secciones interactivas de Google Maps. Por otro lado, Chromedriver se eligió específicamente por su compatibilidad con Google Chrome y su facilidad de integración con Selenium.

El proceso de webcrawling se realizó en la sección de opiniones de usuarios de Google Maps del recinto en estudio. Para ello, se configuró un bot que desplazaba automáticamente la página hacia abajo para cargar todas las opiniones disponibles. Adicionalmente, se programó para hacer clic en el botón "Más"de aquellas opiniones que estaban que no se mostraban completas por su extensión, permitiendo así la recopilación completa de los comentarios presentes en la página. Por último, la calificación dada en cada opinión se extrajo en forma del número de estrellas, con el fin de posteriormente separar las opiniones entre positivas, negativas, y neutras. En total se recopilaron datos de 6 hospitales públicos y 6 clínicas privadas, se recolectaron 1004 opiniones sobre hospitales, y 1037 opiniones sobre clínicas, dando una base de datos total de 2041 opiniones, para lo cual se siguieron ciertos criterios.

4.1.1 Criterios de Selección

Para obtener una muestra lo más representativa posible, se consideraron factores como el tamaño poblacional, la ubicación geográfica, y la cantidad de camas de los recintos seleccionados, esta última característica fue tomada como un determinante de que tan grande era el recinto. Así, se eligieron los hos-

pitales públicos y las clínicas privadas que atienden a una mayor cantidad de población en seis regiones distintas, garantizando un confianza de 95 %, y un error de 0.73 %.

4.1.2 Criterios Éticos

El uso de opiniones de Google Maps no requiere consentimiento informado debido a que el contenido es de acceso público, y los términos y condiciones de la plataforma informan a los usuarios que sus comentarios serán visibles. No obstante, Google es el custodio de los datos, por lo que solo Google y sus API oficiales pueden hacer uso de ellos. En este caso, el uso de Chromedriver, una API de Google, asegura que el proceso de webcrawling no infringe ninguna norma establecida por la plataforma.

Para la recolección de la base de datos, se implementaron las siguientes consideraciones éticas:

1. Los nombres de las personas no fueron incluidos en el análisis.
2. Los nombres de hospitales y clínicas fueron reemplazados por seudónimos (Hospital 1, Hospital 2, Clínica 1, Clínica 2, etc.).
3. No se emplearon frases textuales completas en el reporte; únicamente se presenta información estadística, como frecuencia de palabras y cantidad de opiniones.

4.2 Preprocesamiento

En el contexto de esta investigación, el preprocesamiento del texto fue esencial para asegurar la calidad de los datos antes del análisis. Se aplicaron tres funciones clave: la primera, strip_accents, elimina acentos y caracteres especiales, de modo que el texto quede estandarizado, facilitando la comparación entre palabras y reduciendo el ruido en el análisis. La segunda función, tokenize_text, realiza la tokenización del texto, transformándolo en palabras individuales y eliminando elementos irrelevantes como stopwords en español (por ejemplo, “de”, “el”) y palabras demasiado cortas (menos de dos letras), además de convertir todas las palabras a minúsculas. Por último, se utilizó una función llamada preprocess_and_vectorize, para vectorizar los tokens obtenidos y de esta forma, convertirlos a una representación que el algoritmo sea capaz de interpretar. De esta forma, se logró obtener una representación de los comentarios de los pacientes lista para la posterior extracción de tópicos y análisis de opinión.

4.3 Análisis de Frecuencia y N-gramas

Se realizó un análisis para identificar las palabras y combinaciones de palabras (n-gramas) más frecuentes en los comentarios recolectados.

El procedimiento comenzó con la limpieza y normalización de los textos. Para ello, se eliminó la acentuación mediante la función strip_accents, y posteriormente se aplicó una tokenización que eliminó saltos de línea, transformó todo el texto a minúsculas y descartó las stopwords en español, así como aquellas palabras con menos de dos caracteres.

Una vez preprocesados los textos, se construyó un corpus de tokens tanto para hospitales como para clínicas. A partir de dichos corpus, se contabilizó la frecuencia de aparición de cada palabra, generando una lista con las diez palabras más frecuentes en cada caso. Esta información permitió identificar los términos más representativos en las experiencias de los usuarios.

Además del análisis de frecuencia de palabras individuales, se generaron bigramas y trigramas. Para ello se utilizaron las funciones ngrams de NLTK, adaptadas para crear estas secuencias y calcular su frecuencia de aparición en los textos. De esta manera, fue posible identificar las expresiones más repetidas por los usuarios, capturando frases comunes que reflejan aspectos recurrentes en la percepción del servicio.

4.4 Clasificación de Sentimiento

Para realizar la clasificación de sentimientos en opiniones de usuarios sobre servicios de salud, se utilizó el sistema de calificación adjunto a cada comentario: evaluaciones de 1 o 2 puntos se consideraron negativas, 3 puntos como neutras y 4 o 5 puntos como positivas. Gracias a este criterio, fue posible etiquetar automáticamente los datos sin necesidad de intervención manual, generando un conjunto de datos adecuado para el entrenamiento de modelos de aprendizaje supervisado. En el caso de los hospitales se identificaron 476 opiniones negativas, 468 positivas y 60 neutras; mientras que en las clínicas, se contabilizaron 540 opiniones negativas, 453 positivas y 44 neutras.

Dado que las opiniones ya estaban etiquetadas, se empleó un enfoque supervisado para la clasificación de sentimiento. Se entrenaron y compararon tres modelos: Naive Bayes, SVM y una Red Neuronal Simple. Estos modelos fueron seleccionados por su eficacia reportada en la literatura para tareas de análisis de texto. Los datos fueron previamente vectorizados mediante la técnica TF-IDF para convertir el

texto en una representación numérica interpretable por los algoritmos.

Para evaluar el rendimiento de los modelos, el conjunto de datos fue dividido en un 70 % para entrenamiento y un 30 % para prueba mediante la función train_test_split de sklearn, asegurando una distribución aleatoria de las clases. Posteriormente, se utilizó el informe de clasificación de sklearn para obtener métricas clave como precisión, recall y F1-score. Además, se visualizaron las matrices de confusión y reportes de clasificación mediante gráficos para facilitar la comparación del desempeño entre modelos en la identificación de opiniones positivas, negativas y neutras.

Finalmente, se comprobó que las opiniones neutras representaban una proporción muy baja dentro del conjunto de datos. Por esta razón, se optó por eliminarlas, transformando el problema en una clasificación binaria (positivo vs negativo). Esta decisión buscó mejorar el rendimiento y la capacidad de discriminación de los modelos.

4.5 Clasificación de Tópicos

Para identificar los temas presentes en las opiniones de los usuarios, se implementaron dos enfoques de modelado de tópicos: LDA y FMN. Ambos métodos se aplicaron a los comentarios positivos y negativos por separado, tanto en hospitales como en clínicas, permitiendo explorar diferencias temáticas en función del tipo de establecimiento y la polaridad de la opinión.

El modelo LDA fue entrenado utilizando la biblioteca gensim, con previo preprocesamiento de los textos. Se construyó un diccionario de términos y un corpus en formato bag-of-words para representar los datos. El modelo fue evaluado visualmente mediante nubes de palabras que muestran las palabras más representativas de cada tópico. Para cada conjunto de datos, se realizaron pruebas con diferentes números de tópicos (2, 5 y 7) con el objetivo de identificar la mayor variedad de tópicos presentes en las opiniones de los usuarios.

Como segundo enfoque, se aplicó el modelo FMN a los mismos conjuntos de opiniones, esta vez trabajando directamente sobre la matriz TF-IDF de los textos completos. Los resultados se visualizaron mediante gráficos de barras horizontales, donde se mostraron las diez palabras más importantes de cada tópico. También en este caso se exploraron diferentes números de tópicos (2, 5 y 7) con la misma finalidad.

Capítulo 5. Resultados

5.1 Análisis de Frecuencia y N-gramas

En el análisis de frecuencia y n-gramas, los resultados evidencian que tanto en los comentarios sobre hospitales (Fig.5.2, Fig.5.3 y Fig.5.4) como en los de clínicas (Fig.5.5, Fig.5.6 y Fig.5.7), existe una coexistencia de términos y combinaciones asociadas a experiencias tanto positivas como negativas, con una distribución relativamente equilibrada. En ambos casos destacan palabras como “buena”, “excelente”, “atención”, “mala” y “pésima”, lo que sugiere que los usuarios tienden a manifestar tanto reconocimiento como disconformidad respecto a la calidad del servicio recibido. Este patrón también se refleja en los bigramas más frecuentes, donde aparecen expresiones como “buena atención”, “excelente atención”, “pésima atención” y “mala atención”, así como en los trigramas, que incluyen frases recurrentes como “buena atención personal”, “mala atención personal” o “horas esperando atención”. Por lo tanto, este análisis preliminar muestra que la percepción de la calidad del servicio es mixta tanto en hospitales como en clínicas, sin que se observe un predominio claro de un tono exclusivamente positivo o negativo en alguno de los dos tipos de instituciones.

```
Las 10 palabras más frecuentes son:
Palabra: atencion, Frecuencia: 518
Palabra: hospital, Frecuencia: 383
Palabra: mas, Frecuencia: 241
Palabra: personal, Frecuencia: 229
Palabra: horas, Frecuencia: 166
Palabra: si, Frecuencia: 156
Palabra: buena, Frecuencia: 141
Palabra: excelente, Frecuencia: 129
Palabra: pacientes, Frecuencia: 126
Palabra: gracias, Frecuencia: 122
```

Figura 5.2: Palabras más frecuentes en hospitales.

```
Los 10 bigramas más frecuentes son:
Bigrama: buena atencion, Frecuencia: 71
Bigrama: pesima atencion, Frecuencia: 49
Bigrama: excelente atencion, Frecuencia: 44
Bigrama: mala atencion, Frecuencia: 34
Bigrama: muchas gracias, Frecuencia: 33
Bigrama: salud publica, Frecuencia: 24
Bigrama: atencion personal, Frecuencia: 22
Bigrama: horas esperando, Frecuencia: 19
Bigrama: horas espera, Frecuencia: 18
Bigrama: pesimo servicio, Frecuencia: 16
```

Figura 5.3: Bigramas más frecuentes en hospitales.

```
Los 10 trigramas más frecuentes son:
Trigrama: atendieron super bien, Frecuencia: 4
Trigrama: quiero agradecer personal, Frecuencia: 4
Trigrama: horas esperando atencion, Frecuencia: 4
Trigrama: mas horas esperando, Frecuencia: 3
Trigrama: quiero dar gracias, Frecuencia: 3
Trigrama: mala atencion personal, Frecuencia: 3
Trigrama: llevo horas esperando, Frecuencia: 3
Trigrama: horas media esperando, Frecuencia: 3
Trigrama: buena atencion personal, Frecuencia: 3
Trigrama: pesima atencion parte, Frecuencia: 3
```

Figura 5.4: Trigramas más frecuentes en hospitales.

```
Las 10 palabras más frecuentes son:
Palabra: atencion, Frecuencia: 714
Palabra: clinica, Frecuencia: 463
Palabra: mas, Frecuencia: 338
Palabra: hora, Frecuencia: 302
Palabra: si, Frecuencia: 201
Palabra: medico, Frecuencia: 180
Palabra: buena, Frecuencia: 180
Palabra: personal, Frecuencia: 180
Palabra: excelente, Frecuencia: 178
Palabra: horas, Frecuencia: 170
```

Figura 5.5: Palabras más frecuentes en clínicas.

```
Los 10 bigramas más frecuentes son:
Bigrama: buena atencion, Frecuencia: 83
Bigrama: pesima atencion, Frecuencia: 82
Bigrama: excelente atencion, Frecuencia: 70
Bigrama: mala atencion, Frecuencia: 40
Bigrama: pesimo servicio, Frecuencia: 36
Bigrama: call center, Frecuencia: 25
Bigrama: mala experiencia, Frecuencia: 25
Bigrama: atencion personal, Frecuencia: 24
Bigrama: muchas gracias, Frecuencia: 23
Bigrama: falta respeto, Frecuencia: 22
```

Figura 5.6: Bigramas más frecuentes en clínicas.

```
Los 10 trigramas más frecuentes son:
Trigrama: esperar mas hora, Frecuencia: 9
Trigrama: excelente atencion personal, Frecuencia: 8
Trigrama: buena atencion general, Frecuencia: 6
Trigrama: hacen perder tiempo, Frecuencia: 5
Trigrama: buena atencion personal, Frecuencia: 4
Trigrama: mas dos horas, Frecuencia: 4
Trigrama: buena experiencia clinica, Frecuencia: 4
Trigrama: atencion tiempo espera, Frecuencia: 4
Trigrama: deja bastante desear, Frecuencia: 4
Trigrama: pesima atencion parte, Frecuencia: 4
```

Figura 5.7: Trigramas más frecuentes en clínicas.

5.2 Clasificación Supervisada (Naive Bayes, SVM, Red Neuronal)

En una primera instancia se trabajó considerando 3 clases: positiva, neutra y negativa, sin embargo se observó que la clase neutra presentaba una baja frecuencia, lo que inducía un gran desbalance entre las clases con el consecuente perjuicio en el rendimiento de los clasificadores, por lo que se optó por eliminarla. En el caso de los hospitales, el modelo Naive Bayes muestra una mejora considerable en precisión y distribución de clases tras la eliminación de opiniones neutras (Fig.A.28 vs. Fig.5.8), evidenciado también en el reporte de clasificación respectivo (Fig.A.29 vs. Fig.5.9). De forma similar, el modelo SVM mejora su capacidad de discriminación entre clases al pasar de una matriz con varios errores de clasificación (Fig.A.30) a una más equilibrada (Fig.5.10), lo cual se refleja también en el aumento del F1-score para ambas clases (Fig.A.31 vs. Fig.5.11). El modelo de Red Neuronal, por su parte, presenta la mayor capacidad de ajuste, con una mejora progresiva al simplificar la clasificación binaria (Fig.A.32 vs. Fig.5.12) y métricas significativamente más altas (Fig.A.33 vs. Fig.5.13). En cuanto al rendimiento de cada modelo basado en el accuracy, NB fue el modelo que logró mayor desempeño, seguido del modelo SVM y el modelo de Red Neuronal. El detalle de los valores obtenidos se puede ver en la Tabla.5-1.

En el caso de las clínicas, se repite el patrón de mejora al eliminar la clase neutra. El modelo Naive

Bayes pasa de una clasificación desequilibrada con alta confusión en clase neutra (Fig.A.34) a una mejor segmentación entre positivas y negativas (Fig.5.14), con un reporte de clasificación más consistente (Fig.A.37 vs. Fig.5.15). Para el modelo SVM, la eliminación de la clase intermedia también mejora la separación de clases (Fig.A.36 vs. Fig.5.16), elevando la precisión y recall en ambas etiquetas (Fig.A.37 vs. Fig.5.17). Finalmente, el modelo de Red Neuronal logra un rendimiento sólido, con una clara diagonal dominante en la matriz final (Fig.A.38 vs. Fig.5.18), y métricas balanceadas en su reporte (Fig.A.39 vs. Fig.5.19), lo que confirma su eficacia en entornos con datos más limpios y polarizados. En lo que respecta al rendimiento de cada modelo, SVM fue el modelo que logró mayor desempeño, seguido del modelo de Red Neuronal y el modelo NB. El detalle se puede ver en la Tabla.5-2.

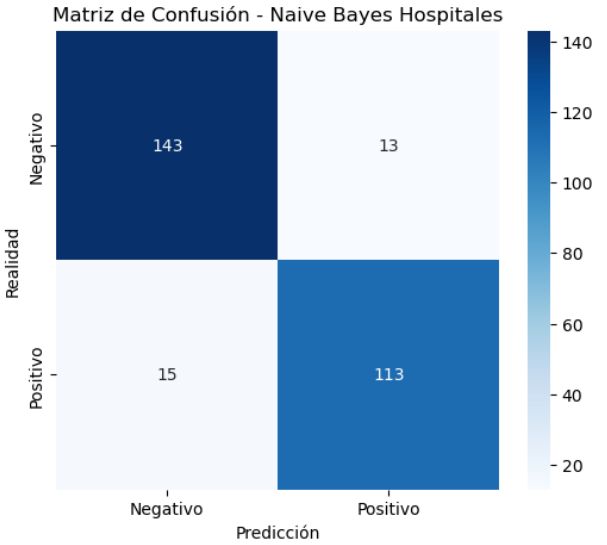


Figura 5.8: Matriz de Confusión NB hospitales considerando 2 clases.

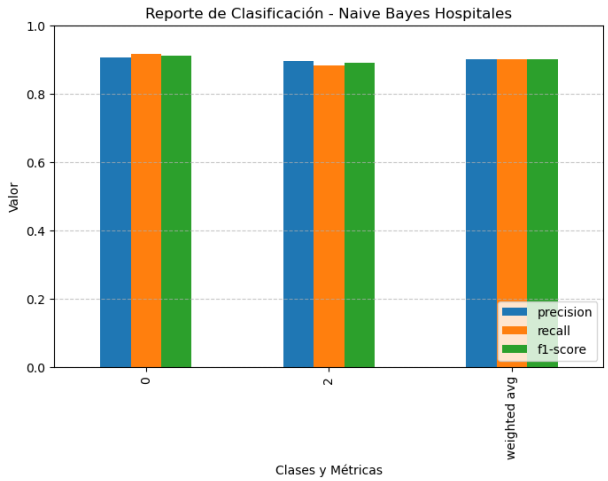


Figura 5.9: Reporte de Clasificación NB hospitales considerando 2 clases.

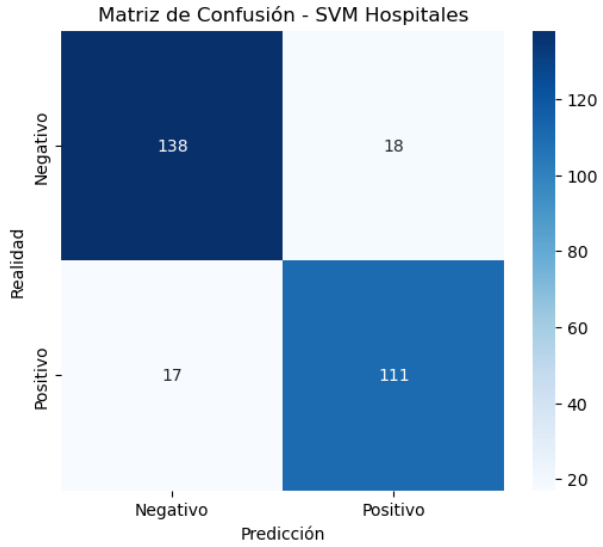


Figura 5.10: Matriz de Confusión SVM hospitales considerando 2 clases.

Figura 5.11: Reporte de Clasificación SVM hospitales considerando 2 clases.

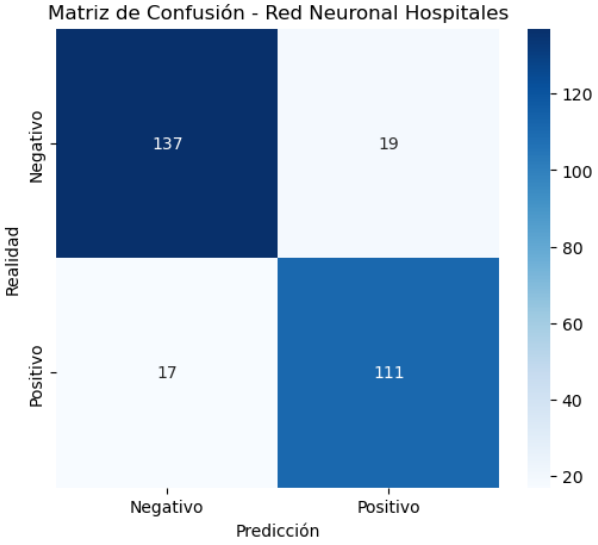


Figura 5.12: Matriz de Confusión Red Neuronal hospitales considerando 2 clases.

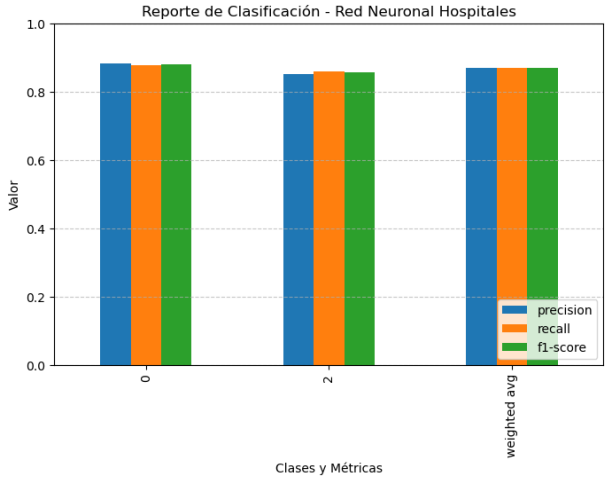


Figura 5.13: Reporte de Clasificación Red Neuronal hospitales considerando 2 clases.

Tabla 5-1: Valores de Accuracy obtenidos para las opiniones de hospitales.

N° de Clases	Naive Bayes	SVM	Red Neuronal
3 Clases	0.8112	0.8112	0.8079
2 Clases	0.9014	0.8767	0.8732

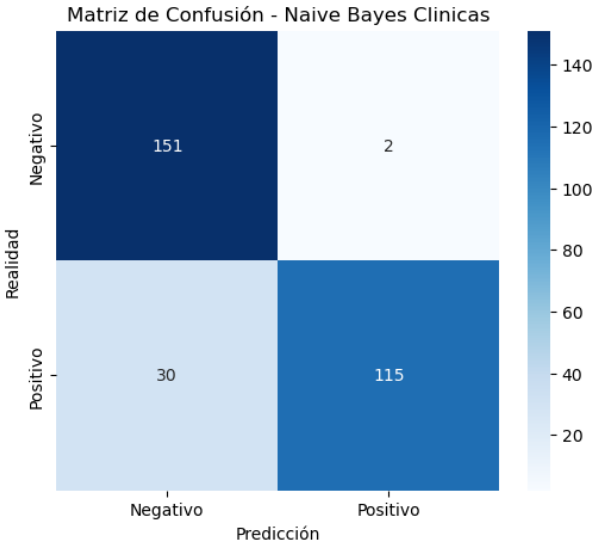


Figura 5.14: Matriz de Confusión NB clinicas considerando 2 clases.

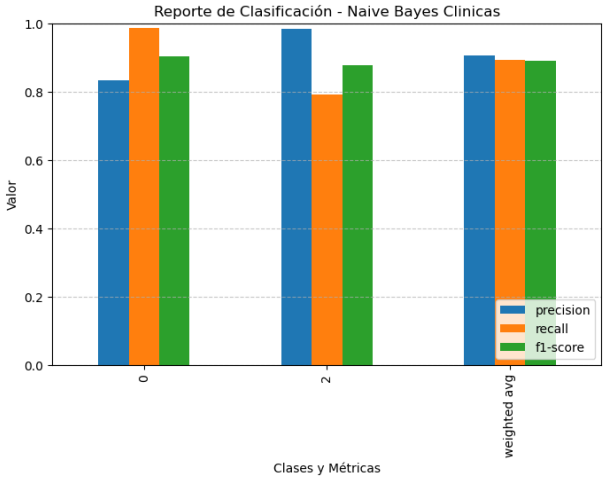


Figura 5.15: Reporte de Clasificación NB clinicas considerando 2 clases.

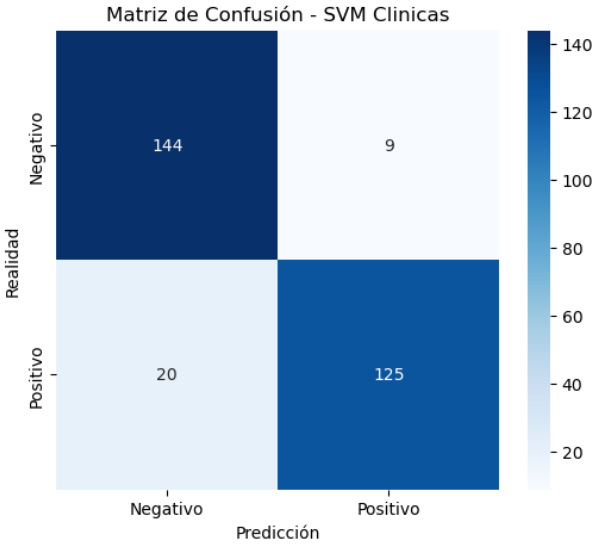


Figura 5.16: Matriz de Confusión SVM clinicas considerando 2 clases.

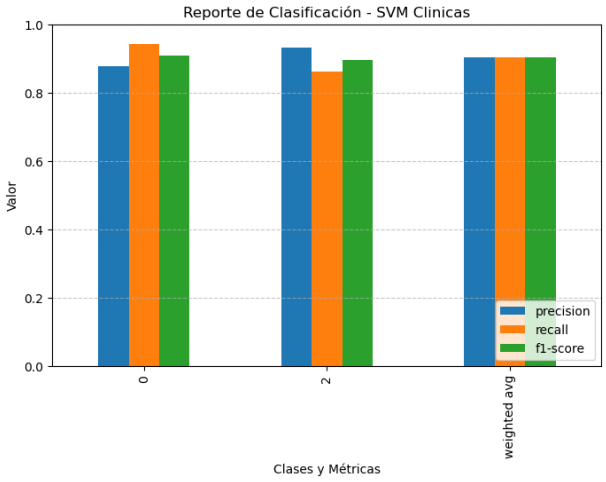


Figura 5.17: Reporte de Clasificación SVM clinicas considerando 2 clases.

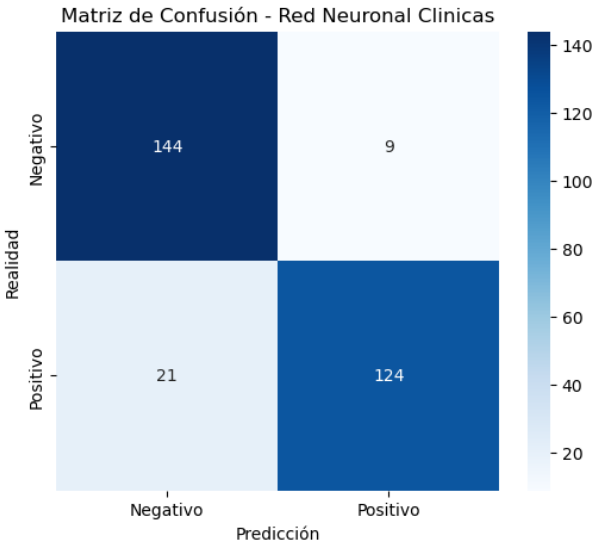


Figura 5.18: Matriz de Confusión Red Neuronal clinicas considerando 2 clases.

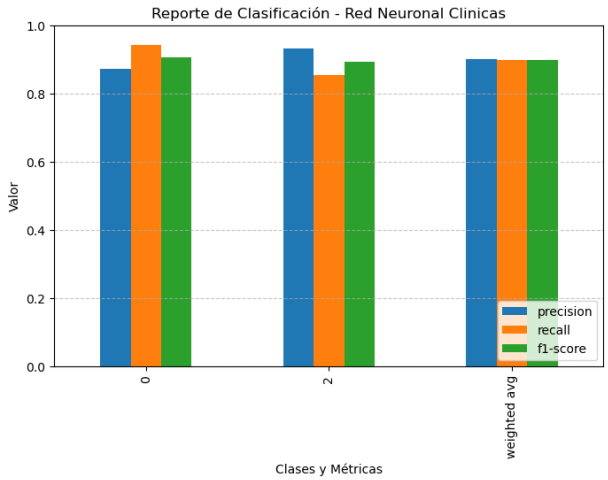


Figura 5.19: Reporte de Clasificación Red Neuronal clinicas considerando 2 clases.

Tabla 5-2: Valores de Accuracy obtenidos para las opiniones de clínicas.

N° de Clases	Naive Bayes	SVM	Red Neuronal
3 Clases	0.8461	0.8653	0.8526
2 Clases	0.8926	0.9026	0.8993

5.3 Modelado de Tópicos (LDA y FMN)

En cuanto al modelado de tópicos, se aplicaron dos enfoques no supervisados: LDA y FMN, tanto en opiniones positivas como negativas, por separado, para hospitales y clínicas. En el caso de LDA aplicado a hospitales, se observa que los tópicos generados permiten identificar patrones claros en opiniones positivas, como son la buena atención general, o la atención en urgencias (Fig.5.20a, Fig.5.20b y Fig.5.20c). A medida que se incrementa el número de tópicos, las nubes de palabras comienzan a mostrar otras categorías, como el buen trato del personal, la buena atención dada a familiares, y agradecimientos de parte de los usuarios. En el caso de las opiniones negativas (Fig.5.21a, Fig.5.21b y Fig.5.21c), los tópicos tienden a centrarse en palabras como “espera”, “urgencia” y “atención”, reflejando temáticas recurrentes como los tiempos de espera y la calidad del trato recibido.

Para las clínicas, los resultados con LDA muestran un comportamiento similar. En las opiniones positivas (Fig.5.22a, Fig.5.22b y Fig.5.22c), los tópicos predominantes aluden a atributos del personal médico y la calidad del servicio (“excelente”, “amable”, “profesionales”), mientras que en las opiniones

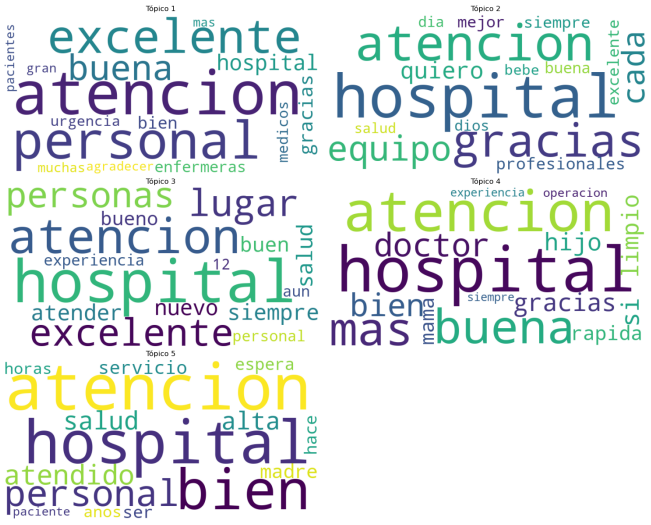
negativas (Fig.5.23a, Fig.5.23b y Fig.5.23c) se mantienen tópicos consistentes con quejas relativas a la espera, atención deficiente y trato. En general, el modelo LDA fue capaz de agrupar palabras con fuerte carga temática, aunque con cierta redundancia cuando se incrementa el número de tópicos.

En el caso del modelo FMN, los resultados se presentan en gráficos de barras horizontales que muestran la importancia de las palabras más representativas de cada tópico. Para hospitales positivos (Fig.5.24a, Fig.5.24b y Fig.5.24c), los temas están bien definidos y reflejan aspectos como profesionalismo, rapidez, el estado de la infraestructura y agradecimiento al personal. En las opiniones negativas de hospitales (Fig.5.25a, Fig.5.25b y Fig.5.25c), FMN permite observar una diferenciación más clara de tópicos que LDA, con agrupaciones centradas en espera, urgencia, mala atención del personal y mala atención telefónica.

En el caso de clínicas, el comportamiento fue similar: en las opiniones positivas (Fig.5.26a, Fig.5.26b y Fig.5.26c), se destacan palabras como “excelente”, “personal”, “trato” y “atención”, mientras que en las negativas (Fig.5.27a, Fig.5.27b y Fig.5.27c), los tópicos enfatizan altos tiempos de espera, malos tratos del personal y problemas en la atención telefónica. A diferencia de LDA, FMN mostró tópicos más consistentes incluso al aumentar el número de temas, manteniendo buena interpretabilidad sin generar redundancia excesiva.

En conjunto, ambos modelos permitieron identificar los temas más recurrentes en las opiniones analizadas. Sin embargo, FMN mostró una mejor separación y claridad en los tópicos generados, especialmente en los subconjuntos de opiniones negativas, lo que sugiere una mayor eficacia del modelo para este tipo de análisis.

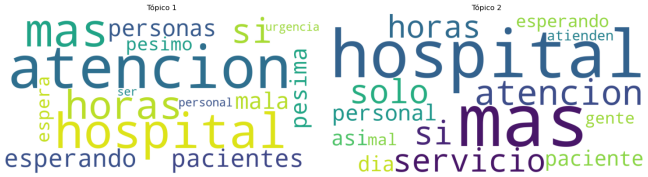
(a) LDA aplicado a hospitales de opinión positiva separado en 2 tópicos.



(b) LDA aplicado a hospitales de opinión positiva separado en 5 tópicos.

(c) LDA aplicado a hospitales de opinión positiva separado en 7 tópicos.

Figura 5.20: Resultados del modelo LDA para hospitales de opinión positiva con distintos números de tópicos: (a) separación en 2 tópicos, (b) separación en 5 tópicos, (c) separación en 7 tópicos.

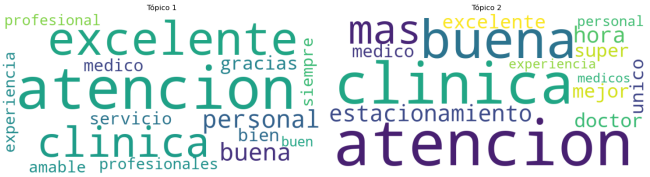


(a) LDA aplicado a hospitales de opinión negativa separado en 2 tópicos.

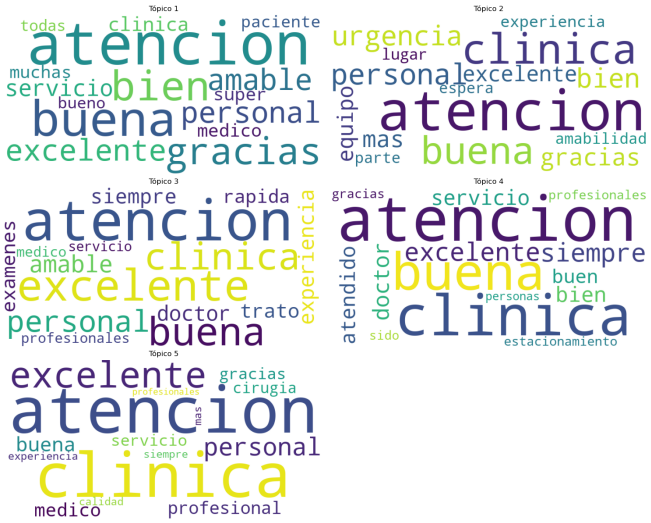
(b) LDA aplicado a hospitales de opinión negativa separado en 5 tópicos.

(c) LDA aplicado a hospitales de opinión negativa separado en 7 tópicos.

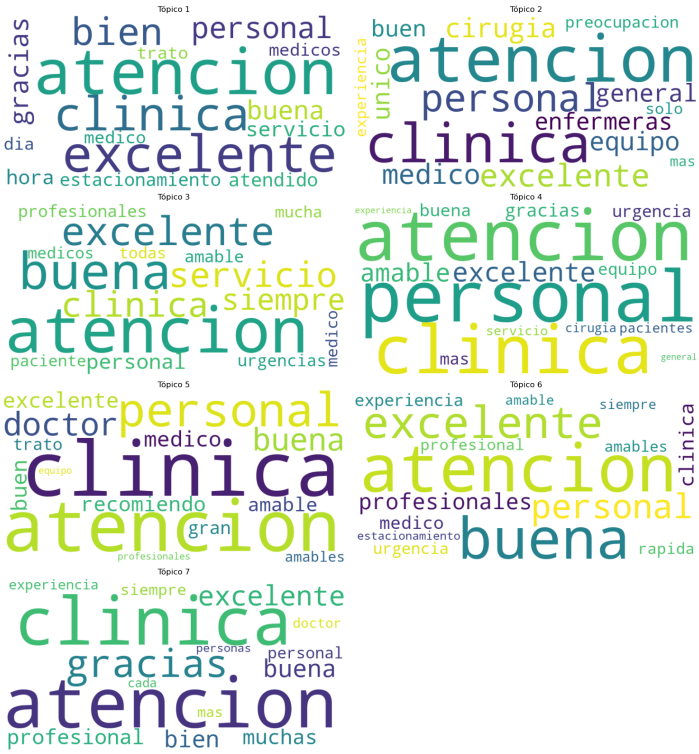
Figura 5.21: Resultados del modelo LDA para hospitales de opinión negativa con distintos números de tópicos: (a) separación en 2 tópicos, (b) separación en 5 tópicos, (c) separación en 7 tópicos.



(a) LDA aplicado a clínicas de opinión positiva separado en 2 tópicos.



(b) LDA aplicado a clínicas de opinión positiva separado en 5 tópicos.

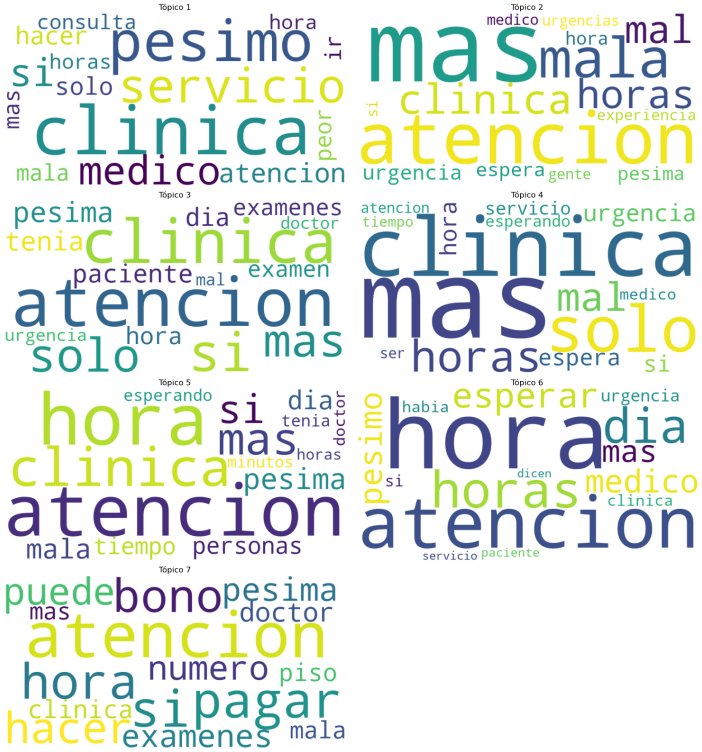


(c) LDA aplicado a clínicas de opinión positiva separado en 7 tópicos.

Figura 5.22: Resultados del modelo LDA para clínicas de opinión positiva con distintos números de tópicos: (a) separación en 2 tópicos, (b) separación en 5 tópicos, (c) separación en 7 tópicos.

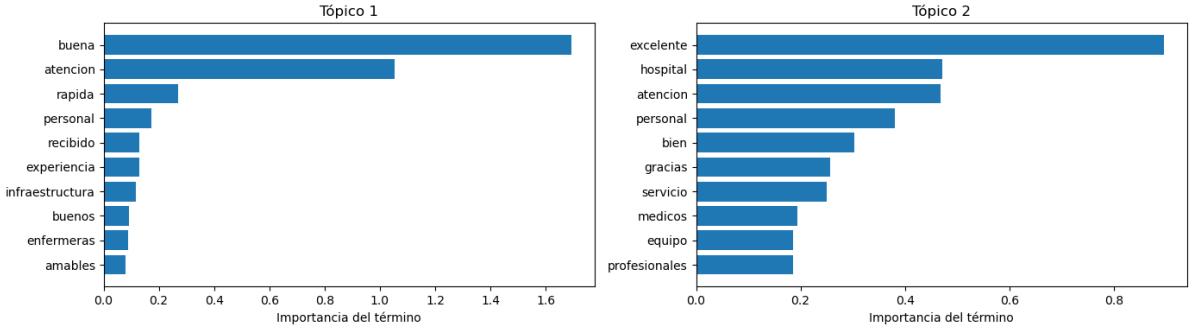
(a) LDA aplicado a clínicas de opinión negativa separado en 2 tópicos.

(b) LDA aplicado a clínicas de opinión negativa separado en 5 tópicos.

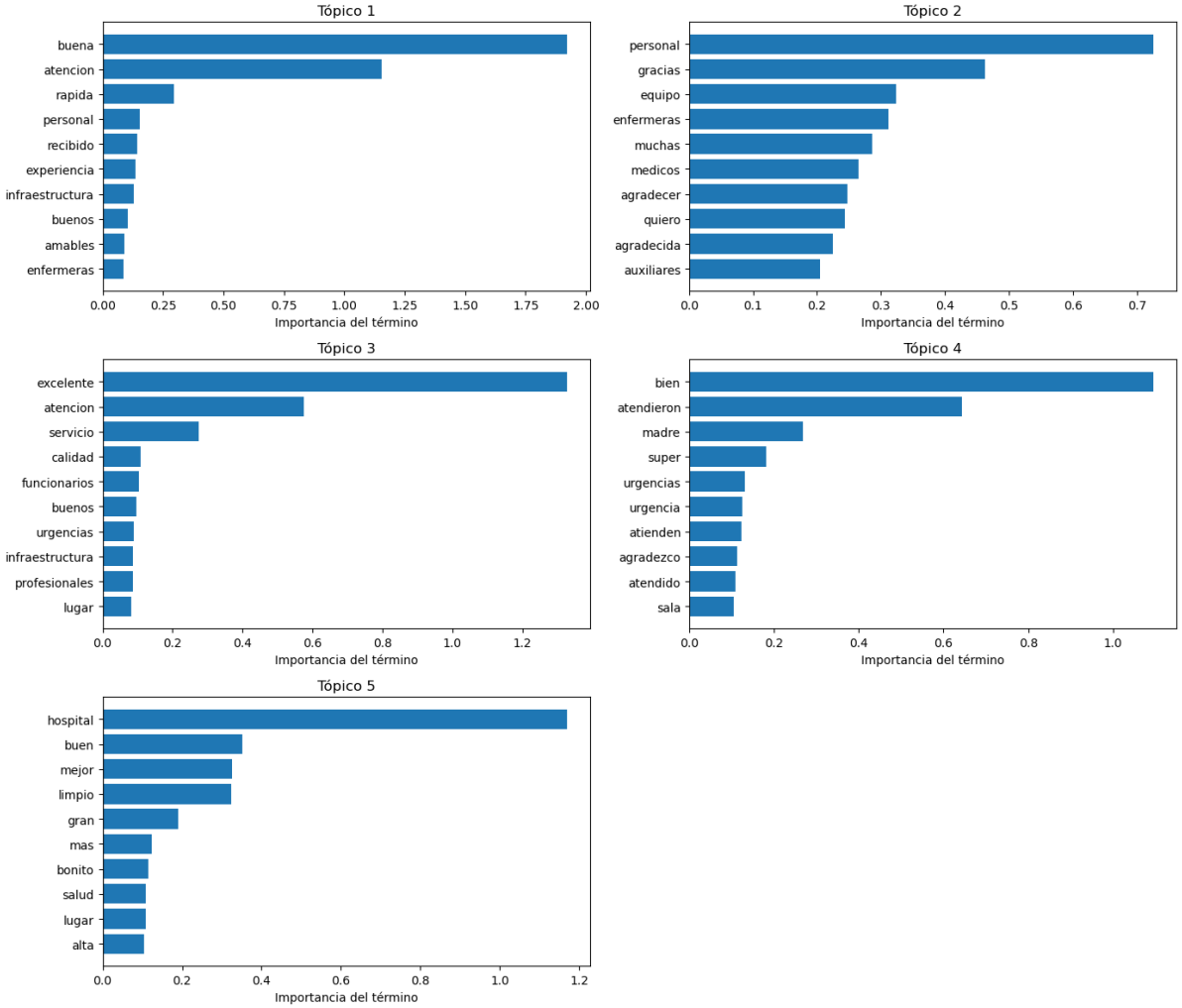


(c) LDA aplicado a clínicas de opinión negativa separado en 7 tópicos.

Figura 5.23: Resultados del modelo LDA para clínicas de opinión negativa con distintos números de tópicos: (a) separación en 2 tópicos, (b) separación en 5 tópicos, (c) separación en 7 tópicos.

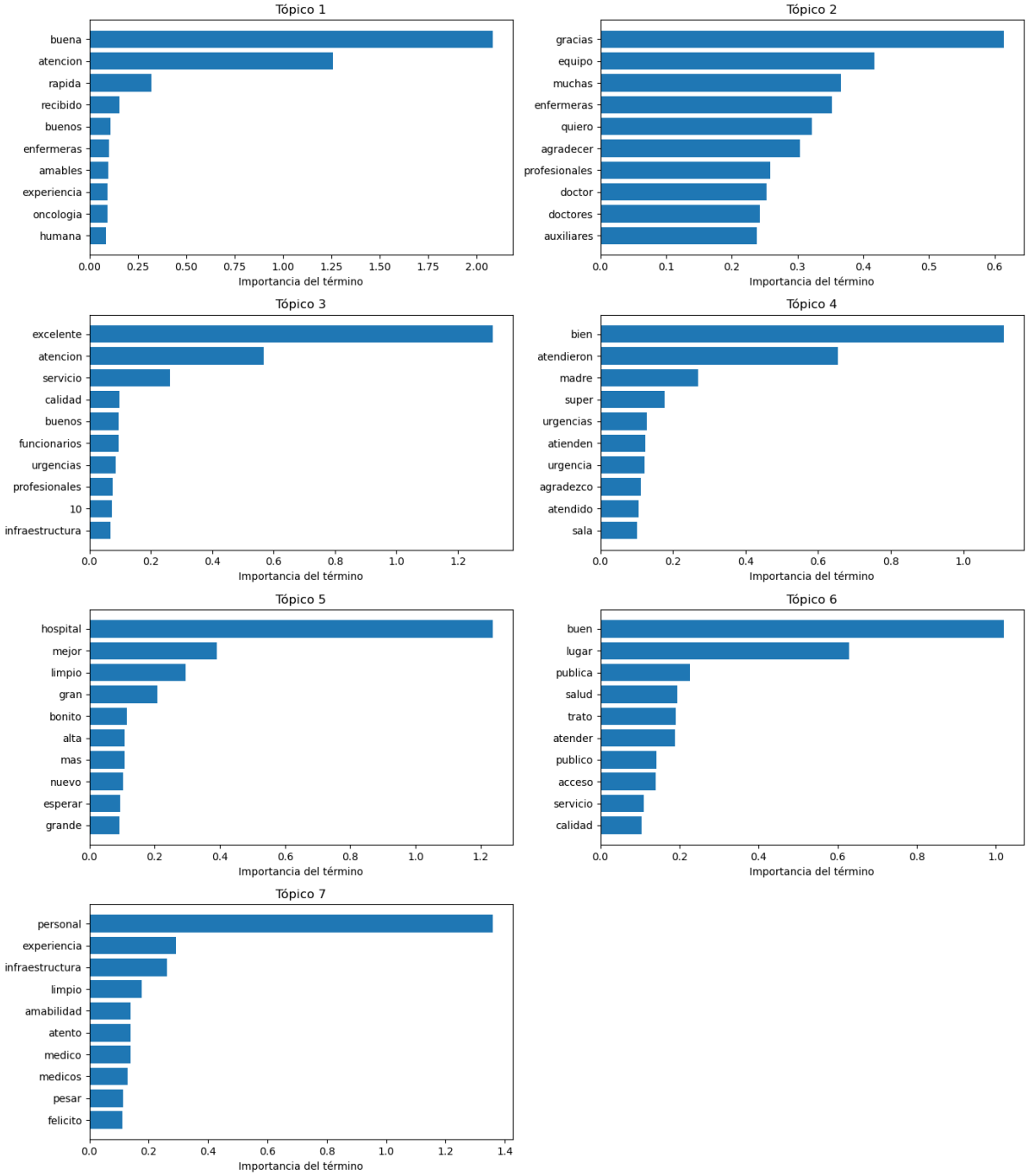


(a) FMN aplicado a hospitales de opinión positiva separado en 2 tópicos.



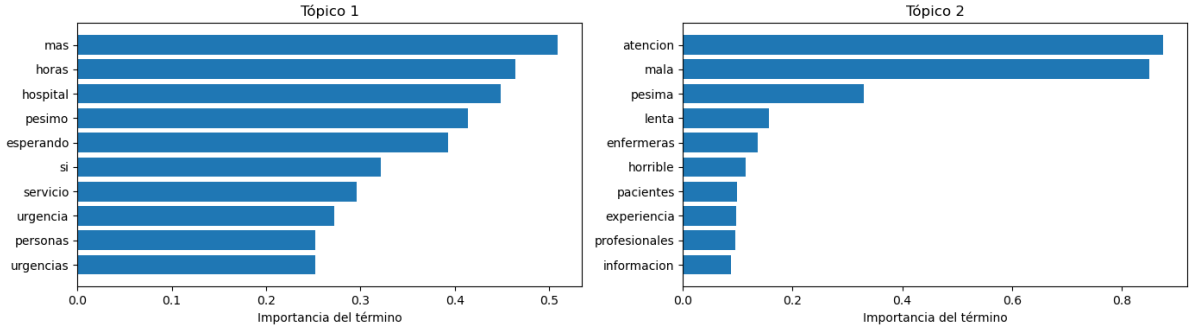
(b) FMN aplicado a hospitales de opinión positiva separado en 5 tópicos.

Figura 5.24: Resultados del modelo FMN para hospitales con opiniones positivas: (a) 2 tópicos, (b) 5 tópicos.

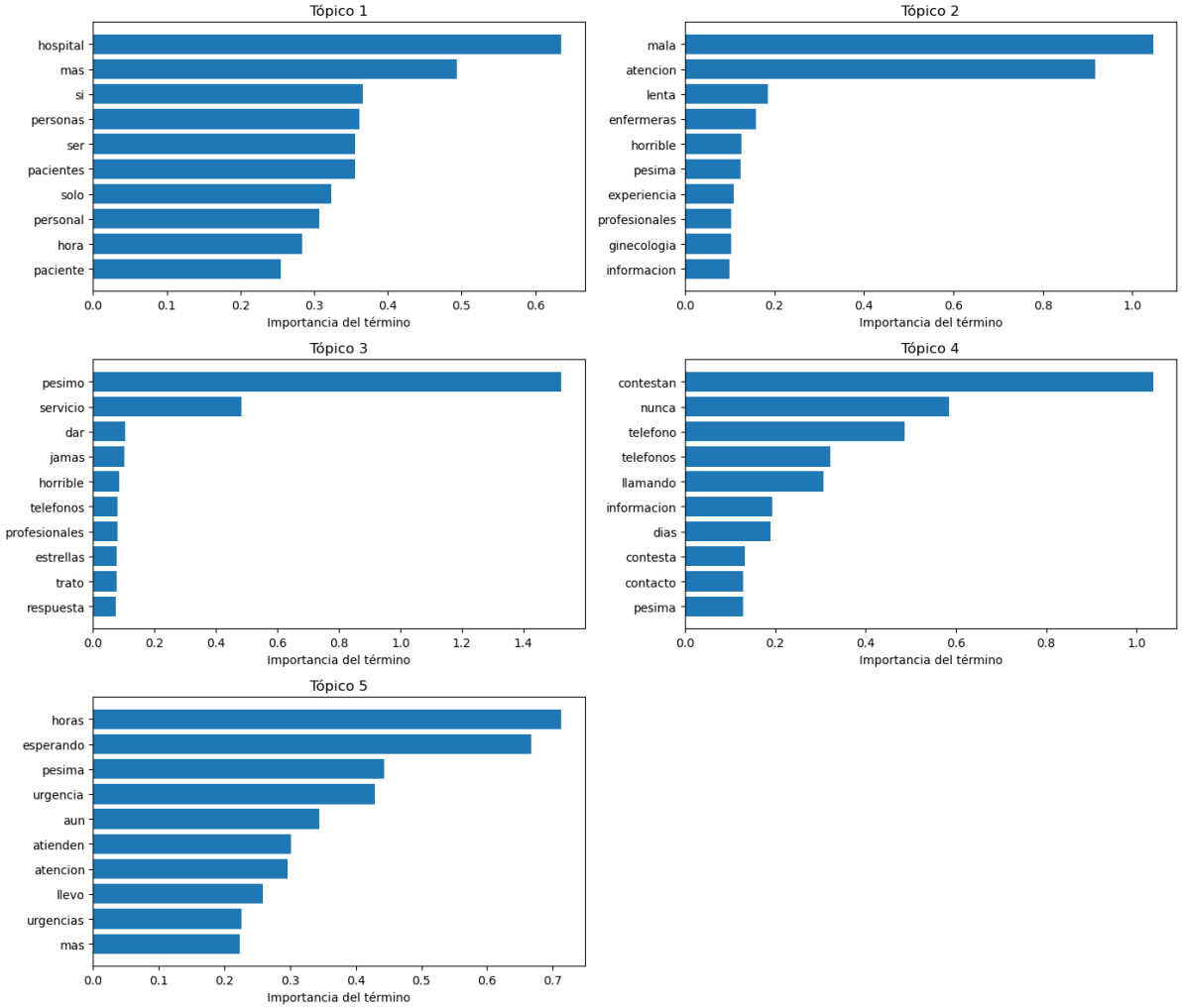


(c) FMN aplicado a hospitales de opinión positiva separado en 7 tópicos.

Continuación de la **Figura 5.24:** Resultados del modelo FMN para hospitales con opiniones positivas: (c) 7 tópicos.

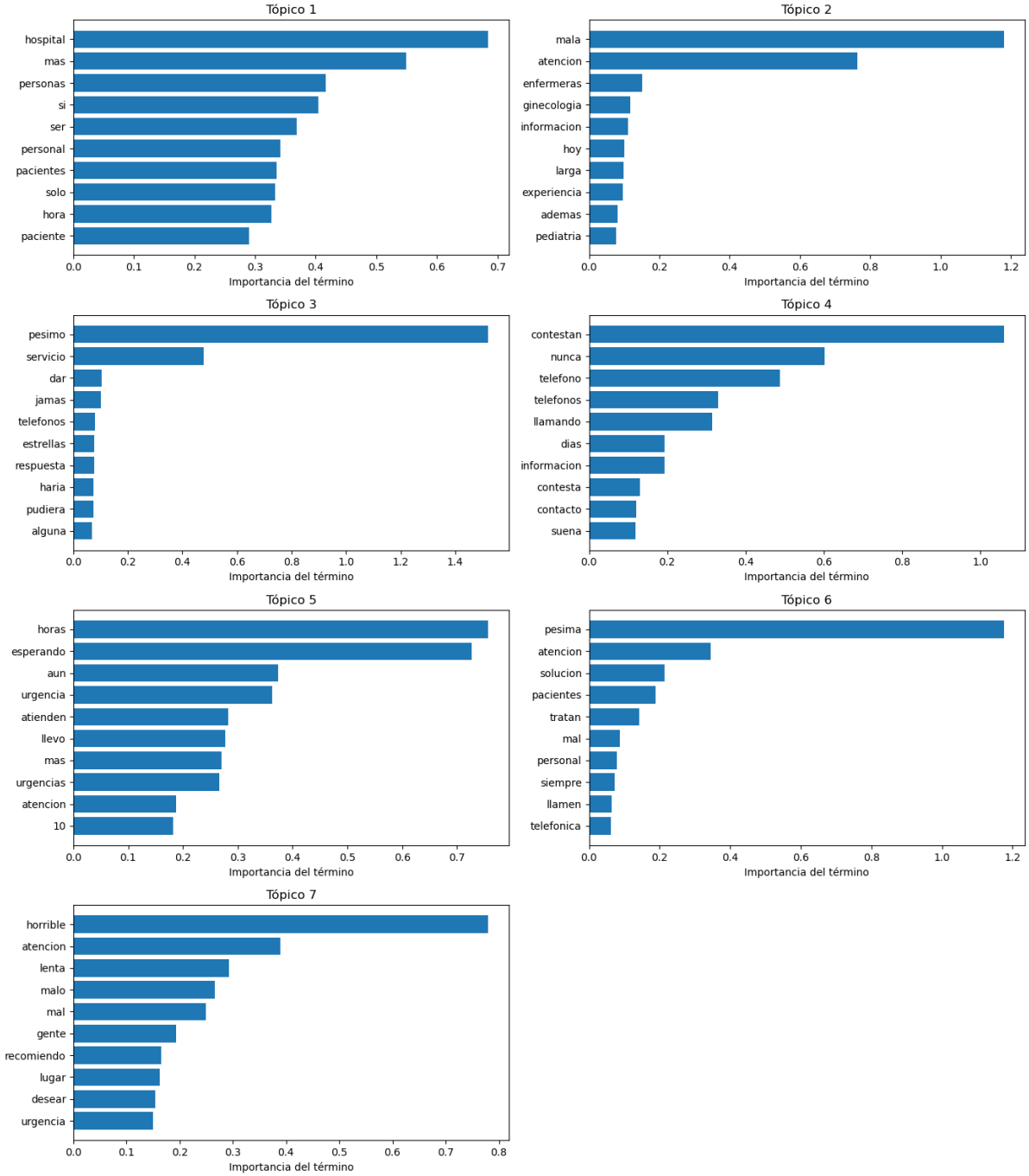


(a) FMN aplicado a hospitales de opinión negativa separado en 2 tópicos.



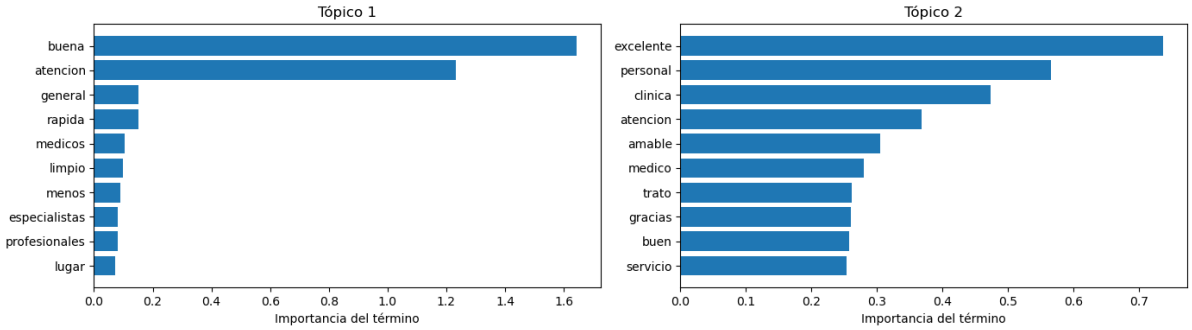
(b) FMN aplicado a hospitales de opinión negativa separado en 5 tópicos.

Figura 5.25: Resultados del modelo FMN para hospitales con opiniones negativas: (a) 2 tópicos, (b) 5 tópicos.

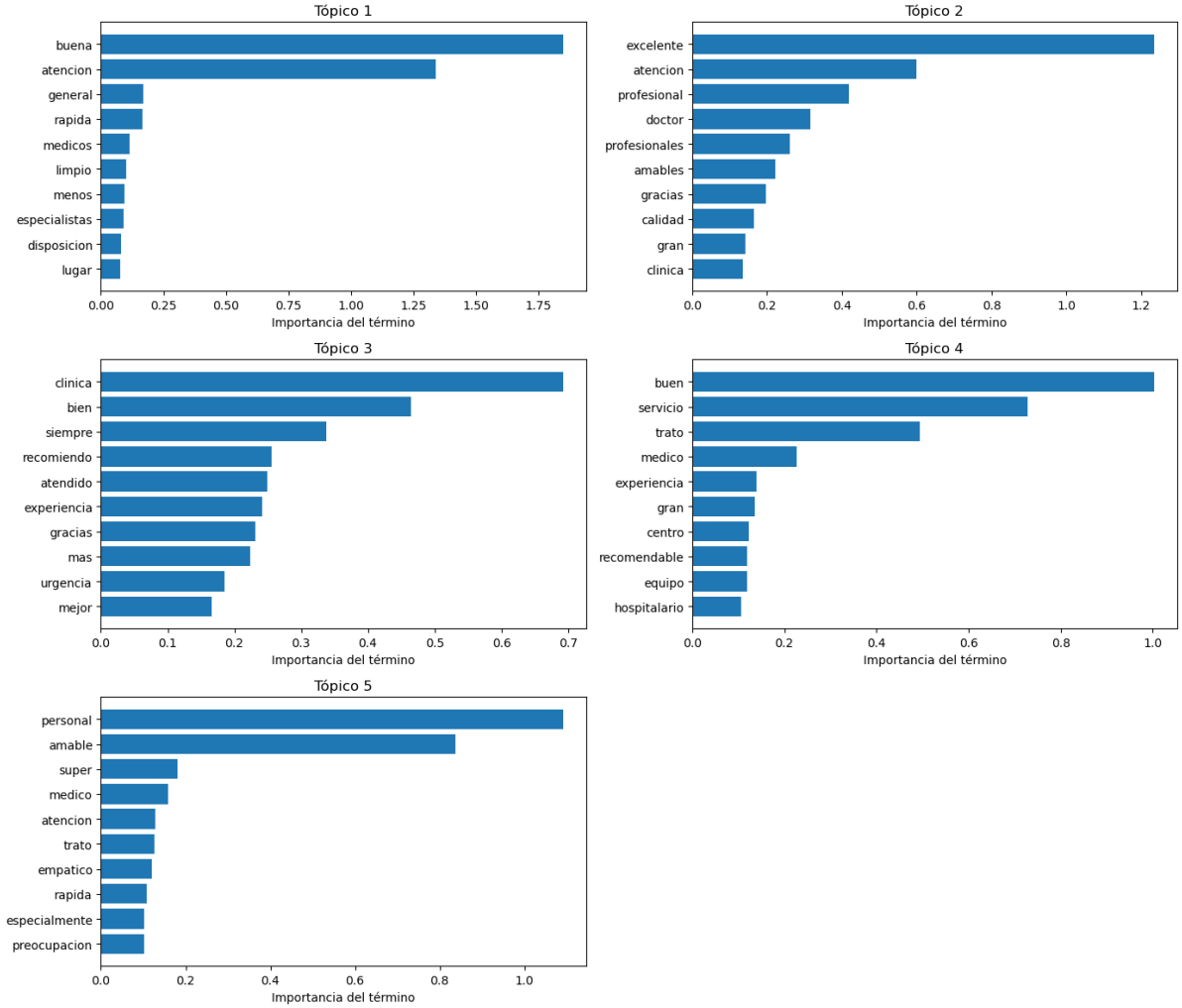


(c) FMN aplicado a hospitales de opinión negativa separado en 7 tópicos.

Continuación de la **Figura 5.24:** Resultados del modelo FMN para hospitales con opiniones negativas: (c) 7 tópicos.

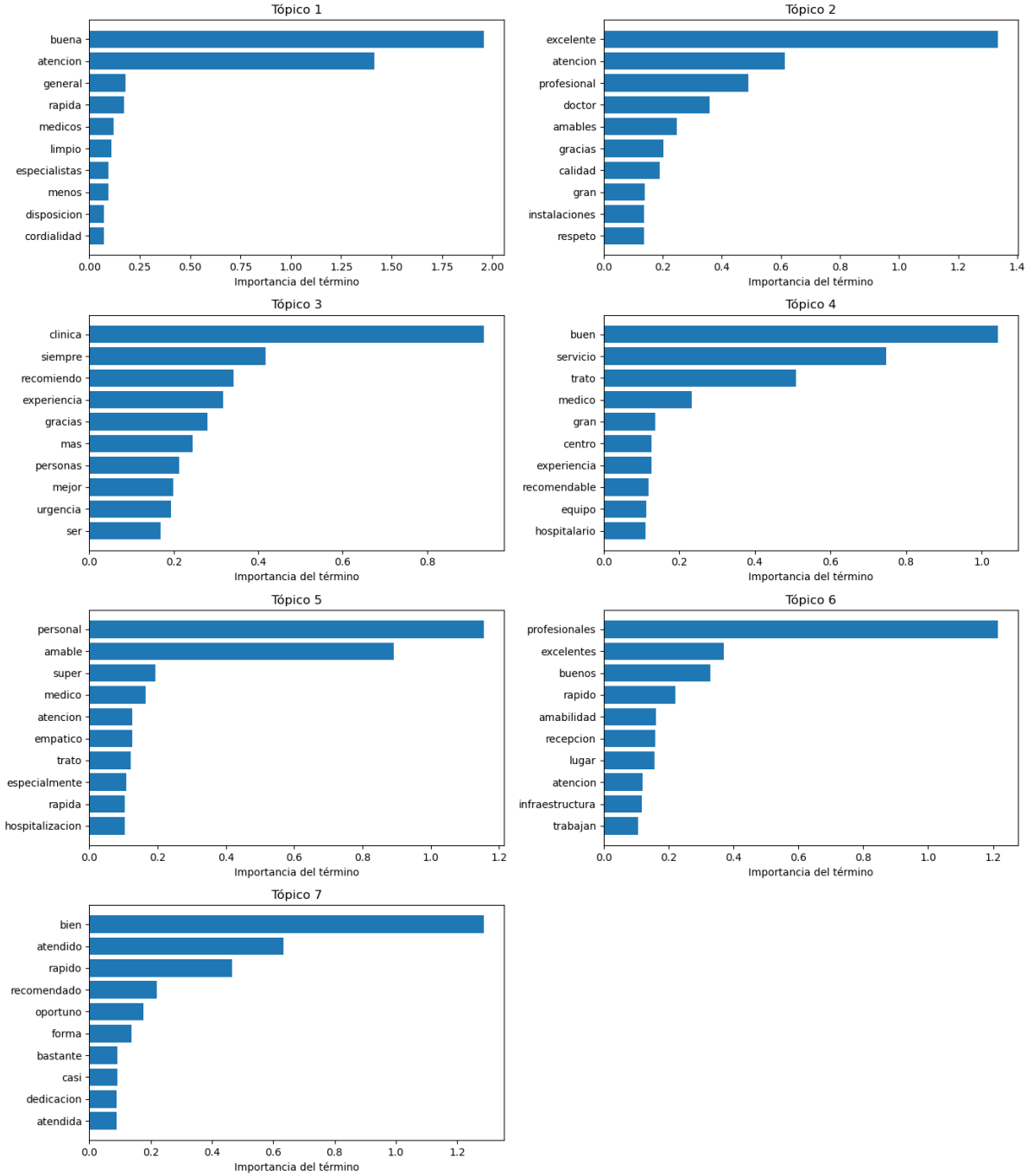


(a) FMN aplicado a clínicas de opinión positiva separado en 2 tópicos.



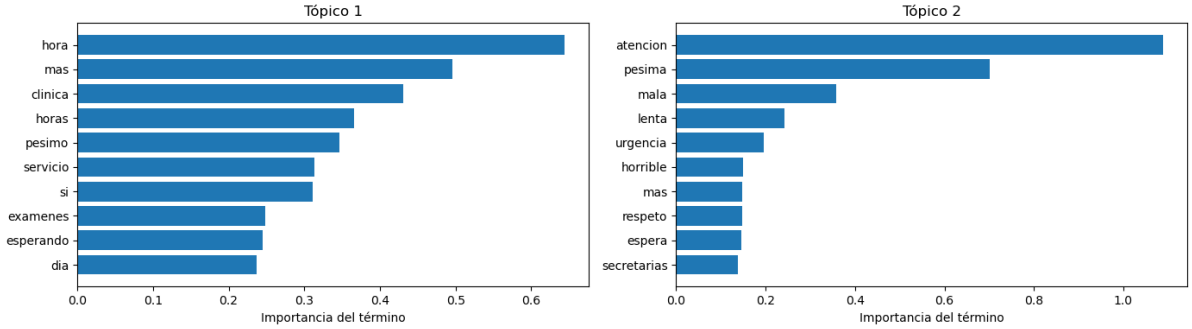
(b) FMN aplicado a clínicas de opinión positiva separado en 5 tópicos.

Figura 5.26: Resultados del modelo FMN para clínicas con opiniones positivas: (a) 2 tópicos, (b) 5 tópicos.

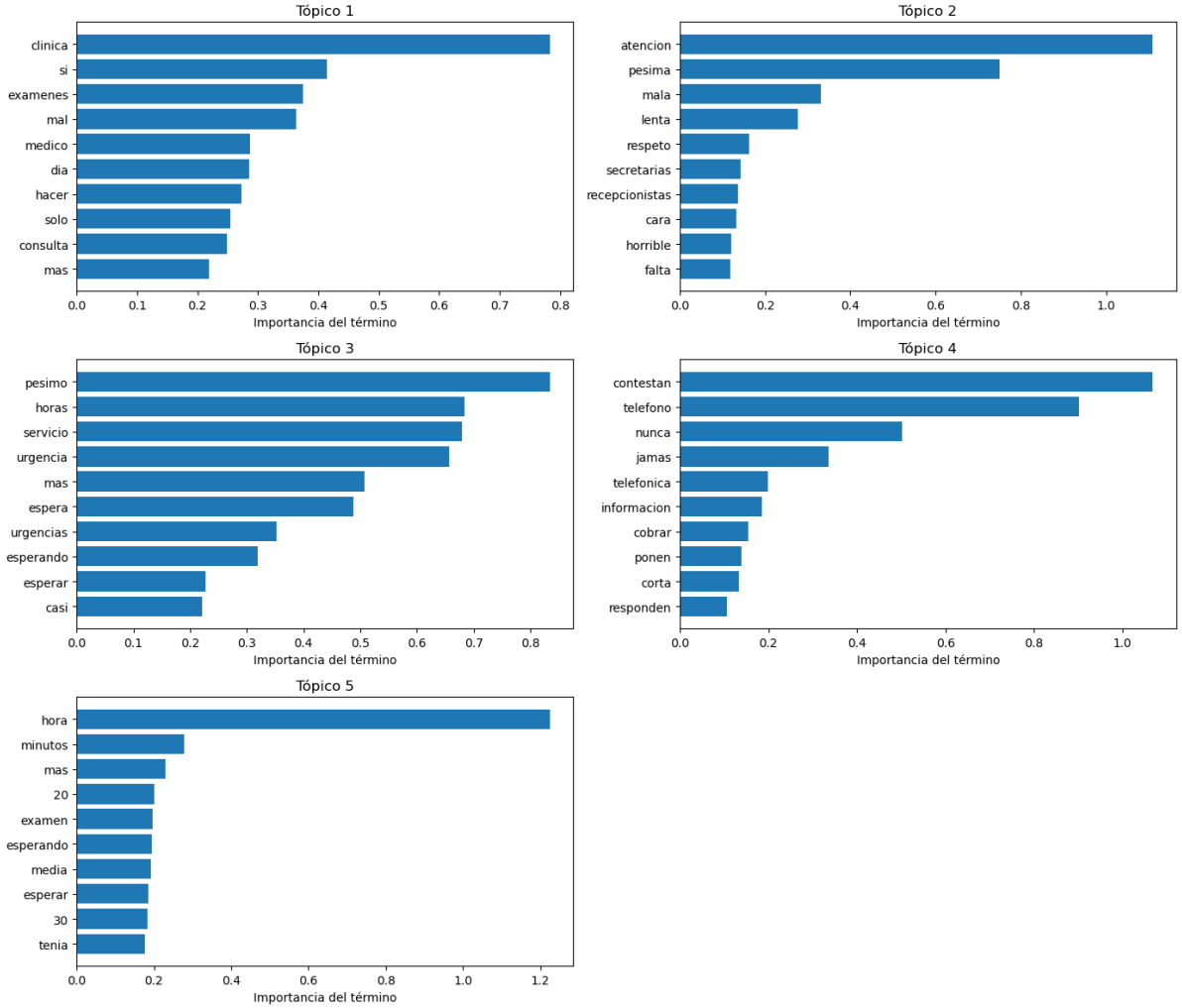


(c) FMN aplicado a clínicas de opinión positiva separado en 7 tópicos.

Continuación de la **Figura 5.26:** Resultados del modelo FMN para clínicas con opiniones positivas: (c) 7 tópicos.

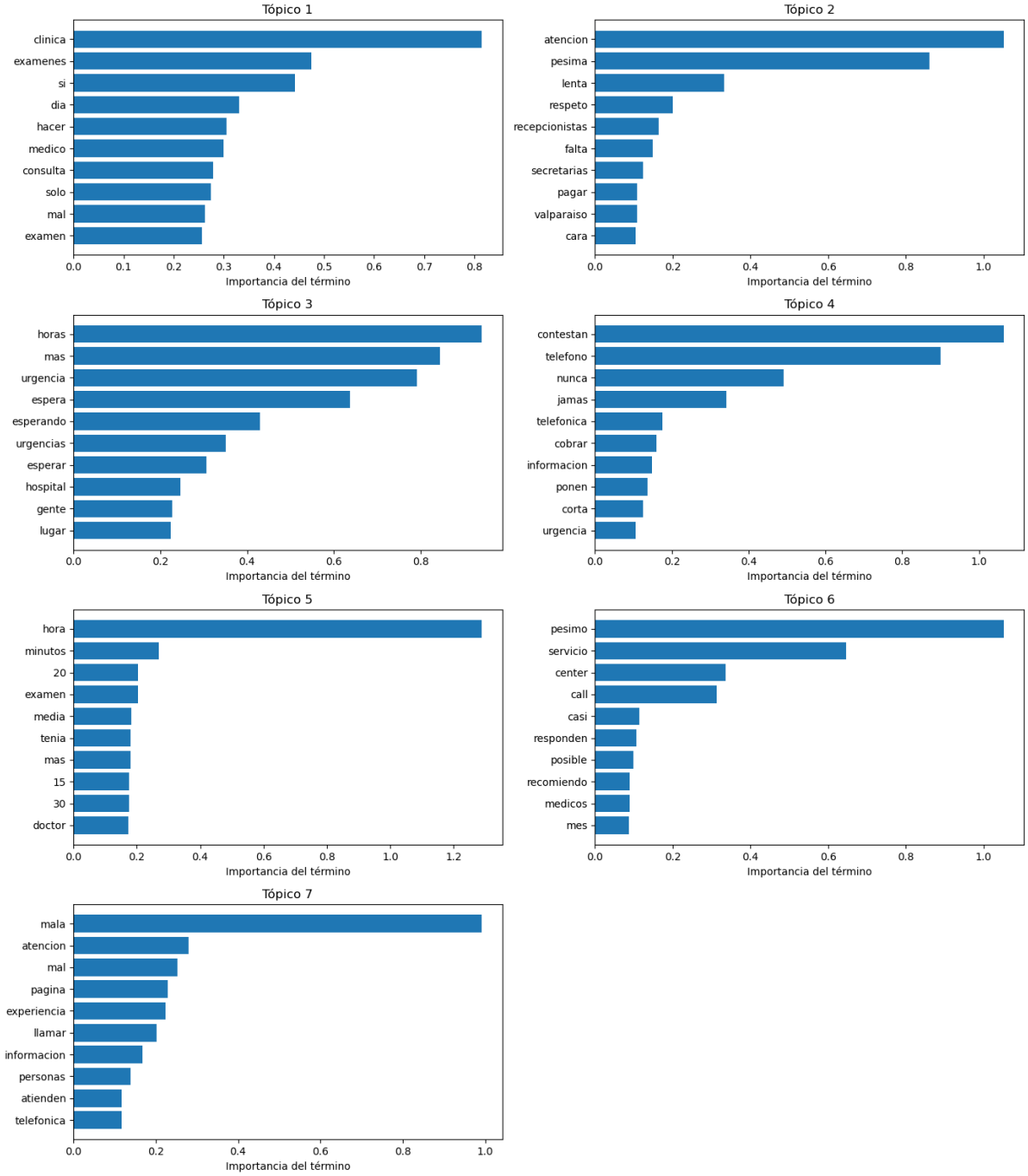


(a) FMN aplicado a clínicas de opinión negativa separado en 2 tópicos.



(b) FMN aplicado a clínicas de opinión negativa separado en 5 tópicos.

Figura 5.27: Resultados del modelo FMN para clínicas con opiniones negativas: (a) 2 tópicos, (b) 5 tópicos.



(c) FMN aplicado a clínicas de opinión negativa separado en 7 tópicos.

Continuación de la **Figura 5.27:** Resultados del modelo FMN para clínicas con opiniones negativas: (c) 7 tópicos.

Capítulo 6. Conclusión

6.1 Discusión

El presente trabajo permitió comprobar la viabilidad de aplicar técnicas de Procesamiento de Lenguaje Natural (PLN) y aprendizaje automático para analizar la percepción de los usuarios sobre la calidad del servicio de salud en Chile, a partir de opiniones públicas recolectadas desde Google Maps. Uno de los principales hallazgos es que, tanto en hospitales públicos como en clínicas privadas, las opiniones presentan una mezcla equilibrada de experiencias positivas y negativas, lo cual se evidenció desde el análisis de frecuencia y n-gramas, y fue confirmado posteriormente mediante modelos de clasificación y modelado de tópicos.

En lo que respecta a la clasificación de sentimientos, los modelos supervisados Naive Bayes, SVM y Red Neuronal fueron capaces de identificar con éxito la polaridad de las opiniones. La eliminación de la clase neutra, justificada por su baja representación, resultó beneficiosa para mejorar el desempeño de los modelos. Si bien todos mostraron desempeños aceptables, los resultados evidenciaron que la efectividad de cada modelo varió según el tipo de institución analizada. Por ejemplo, Naive Bayes presentó el mayor desempeño en hospitales, mientras que SVM lo hizo en clínicas, lo cual sugiere que las características del lenguaje usado por los usuarios pueden incidir en la capacidad predictiva de cada algoritmo. Sobre el modelo de Red Neuronal en hospitales, este fue el modelo con menor desempeño, alineándose con lo señalado en la bibliografía respecto a su bajo rendimiento frente a bases de datos pequeñas. Sin embargo, en el caso de las clínicas, el modelo neuronal superó a Naive Bayes, posicionándose como el segundo mejor clasificador.

Respecto al modelado de tópicos, tanto LDA como FMN permitieron extraer temáticas relevantes en los comentarios. LDA mostró ser útil para un análisis general de temas, pero tendió a generar redundancia en los tópicos al aumentar su número. En contraste, FMN ofreció una segmentación más clara y con mejor interpretabilidad, especialmente en los subconjuntos de opiniones negativas, donde se identificaron tópicos concretos como “larga espera”, “mal trato del personal” o “deficiencia telefónica”. Esta capacidad de distinguir tópicos de forma precisa evidencia el potencial de FMN para su uso en análisis más específicos y dirigidos a la mejora de servicios.

6.2 Conclusión

En base a la investigación realizada y los resultados obtenidos, este trabajo logró demostrar que es posible utilizar técnicas de Procesamiento de Lenguaje Natural (PLN) y aprendizaje automático para analizar de forma automatizada la percepción de los usuarios del sistema de salud chileno a partir de opiniones públicas extraídas desde la web. A través de un enfoque combinado de aprendizaje supervisado y no supervisado, se logró clasificar correctamente el sentimiento de más de 2.000 opiniones y descubrir tópicos latentes relevantes, sin necesidad de intervención manual.

Los modelos supervisados Naive Bayes, SVM y Red Neuronal lograron desempeños aceptables en la tarea de clasificación de sentimientos, siendo el modelo Naive Bayes el más eficaz en hospitales y SVM en clínicas. En particular, se observó que la red neuronal, a pesar de haber sido el modelo con menor precisión en hospitales —tal como predecía la literatura para bases de datos pequeñas—, superó al modelo Naive Bayes en clínicas, mostrando que su rendimiento puede mejorar en contextos con una estructura de lenguaje más clara. Por otra parte, la eliminación de la clase neutra resultó ser una decisión acertada, ya que permitió simplificar el problema a una clasificación binaria, lo que aumentó la precisión de todos los modelos.

En cuanto al modelado de tópicos, los enfoques no supervisados Latent Dirichlet Allocation (LDA) y Factorización de Matrices No Negativas (FMN) permitieron identificar las temáticas más recurrentes en los comentarios, como la calidad de la atención, el trato del personal, la espera y la organización. LDA resultó útil para una exploración temática general, mientras que FMN ofreció una mejor definición de tópicos, especialmente en los subconjuntos negativos, evidenciando su mayor eficacia para la interpretación de patrones en los datos.

El algoritmo desarrollado, además de complementar los métodos tradicionales de análisis, posee el potencial de ser implementado en sistemas de monitoreo de opinión en tiempo real, permitiendo una clasificación inmediata de la polaridad de los comentarios y una interpretación clara de los temas planteados por los usuarios. Esto facilitaría una respuesta oportuna ante problemas emergentes, apoyando así la mejora continua en los servicios de salud en Chile.

Además, el análisis realizado permitió identificar tanto los aspectos más valorados como aquellos que generan mayor insatisfacción entre los pacientes, lo cual constituye una base para que las instituciones fortalezcan lo que funciona bien y enfoquen sus esfuerzos en resolver los problemas más comunes. Sin embargo, si queremos hablar realmente de soluciones estructurales a estos problemas, entonces debemos entrar en el terreno de la política pública, un área que se aleja del alcance de esta investigación.

6.3 Trabajo Futuro

Si bien los resultados obtenidos fueron satisfactorios, existen diversas líneas de trabajo que podrían ser exploradas para ampliar y perfeccionar el análisis realizado. Una primera mejora sería la incorporación de una base de datos de mayor tamaño y diversidad, incluyendo opiniones de más instituciones de salud y considerando otras fuentes públicas como redes sociales o sitios web de reseñas específicas del sector, como Doctoralia. Esto permitiría mejorar los modelos supervisados e implementar arquitecturas de redes neuronales más complejas, que podrían brindar mejores resultados.

Asimismo, sería interesante explorar técnicas avanzadas de Procesamiento de Lenguaje Natural, como modelos preentrenados de lenguaje basados en arquitecturas Transformer, que han demostrado un buen desempeño en tareas de clasificación de sentimientos y extracción de tópicos en corpus más grandes. Esto permitiría comparar su rendimiento con los algoritmos tradicionales utilizados en este trabajo, y evaluar su aplicabilidad en bases de datos de opinión sobre salud.

Otra línea de trabajo futuro corresponde a la implementación de un sistema de clasificación en tiempo real, capaz de integrarse directamente con plataformas de opinión en línea. Esto permitiría a las instituciones de salud realizar un monitoreo continuo de las opiniones emitidas por los usuarios, generando alertas ante tendencias negativas o cambios en la percepción de los pacientes.

Asimismo, se podrían desarrollar sistemas de visualización interactiva de los resultados obtenidos que permitan explorar los tópicos identificados, analizar la evolución del sentimiento de los usuarios y detectar áreas críticas de mejora. Estas herramientas visuales facilitarían la toma de decisiones basada en datos y harían accesible la información a profesionales de la salud sin necesidad de conocimientos técnicos avanzados.

Glosario

FFT Friend and Family Test

FMN Factorización de Matrices No Negativas

FONASA Fondo Nacional de Salud

HCAHPS Hospital Consumer Assessment of Healthcare Providers and Systems

ISAPRE Instituciones de Salud Previsional

LDA Latent Dirichlet Allocation

MESU Medición de Satisfacción Usuaria

NB Naive Bayes

OECD Organización para la Cooperación y el Desarrollo Económico

PLN Procesamiento de Lenguaje Natural

PREM Patient-Reported Experience Measures

RNA Redes Neuronales Artificiales

RNC Redes Neuronales Convergentes

RNR Redes Neuronales Recurrentes

SERVQUAL Service Quality

SVM Máquina de Vectores de Soporte

TF-IDF Term Frequency-Inverse Document Frequency

Referencias

[1] A. Goic, "El sistema de salud de chile: una tarea pendiente," *Revista médica de Chile*, vol. 143, no. 6, pp. 774–786, 2015.

[2] Biblioteca del Congreso Nacional, "Listas y tiempos de espera para atención en salud en Chile," 2024. [Online]. Available: https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/36366/2/BCN_Tiempos_de_espera_para_atencion_en_salud__EG_final.pdf

[3] Superintendencia de Salud, "Boletín Estadístico Informativo IP – Diciembre 2023," 2023. [Online]. Available: https://www.superdesalud.gob.cl/biblioteca-digital/boletin-estadistico-informativo-ip-diciembre-2023/

[4] OECD, "OECD Health Statistics," 2024, accessed: 2024-10-03. [Online]. Available: https://www.oecd.org/en/data/datasets/oecd-health-statistics.html

[5] G. Hu, X. Han, H. Zhou, and Y. Liu, "Public perception on healthcare services: evidence from social media platforms in china," *International journal of environmental research and public health*, vol. 16, no. 7, p. 1273, 2019.

[6] Ministerio de Hacienda. (2024) Medición de Satisfacción Usuaria (MESU). [Online]. Available: https://satisfaccion.gob.cl/medicion-de-satisfaccion-usuaria/proceso-2024

[7] K. Nawab, G. Ramsey, and R. Schreiber, "Natural language processing to extract meaningful information from patient experience feedback," *Applied Clinical Informatics*, vol. 11, no. 02, pp. 242–252, 2020.

[8] E. Vashishtha and H. Kapoor, "Enhancing patient experience by automating and transforming free text into actionable consumer insights: a natural language processing (nlp) approach," *International Journal of Health Sciences and Research*, vol. 13, no. 10, pp. 275–288, 2023.

[9] N. Zaman, D. M. Goldberg, A. S. Abrahams, and R. A. Essig, "Facebook hospital reviews: Automated service quality detection and relationships with patient satisfaction," *Decision Sciences*, vol. 52, no. 6, pp. 1403–1431, 2021.

[10] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, L. Donaldson *et al.*, "Use of sentiment analysis for capturing patient experience from free-text comments posted online," *Journal of medical Internet research*, vol. 15, no. 11, p. e2721, 2013.

[11] S. A. Cammel, M. S. De Vos, D. van Soest, K. M. Hettne, F. Boer, E. W. Steyerberg, and H. Boosman, "How to automatically turn patient experience free-text responses into actionable insights: a natural language programming (nlp) approach," *BMC medical informatics and decision making*, vol. 20, pp. 1–10, 2020.

[12] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi, and E. Mayer, "Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review," *BMJ Health & Care Informatics*, vol. 28, no. 1, 2021.

[13] M. M. van Buchem, O. M. Neve, I. M. Kant, E. W. Steyerberg, H. Boosman, and E. F. Hensen, "Analyzing patient experiences using natural language processing: development and validation of the artificial intelligence patient reported experience measure (ai-prem)," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, p. 183, 2022.

[14] A. I. A. Rahim, M. I. Ibrahim, K. I. Musa, S.-L. Chua, and N. M. Yaacob, "Assessing patient-perceived hospital service quality and sentiment in malaysian public hospitals using machine learning and facebook reviews," *International journal of environmental research and public health*, vol. 18, no. 18, p. 9912, 2021.

[15] G. Shaw Jr, M. Zimmerman, L. Vasquez-Huot, and A. Karami, "Deciphering latent health information in social media using a mixed-methods design," in *Healthcare*, vol. 10, no. 11. MDPI, 2022, p. 2320.

[16] E. Ötleş, D. E. Kendrick, Q. P. Solano, M. Schuller, S. L. Ahle, M. H. Eskender, E. Carnes, and B. C. George, "Using natural language processing to automatically assess feedback quality: findings from 3 surgical residencies," *Academic Medicine*, vol. 96, no. 10, pp. 1457–1460, 2021.

[17] C. J. Harrison and C. J. Sidey-Gibbons, "Machine learning in medicine: a practical introduction to natural language processing," *BMC medical research methodology*, vol. 21, no. 1, p. 158, 2021.

[18] A. M. Shah, X. Yan, S. A. A. Shah, and G. Mamirkulova, "Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 2925–2942, 2020.

[19] P. Lavanya and E. Sasikala, "Deep learning techniques on text classification using natural language processing (nlp) in social healthcare network: A comprehensive survey," in *2021 3rd international conference on signal processing and communication (ICPSC)*. IEEE, 2021, pp. 603–609.

[20] G. B. Silva, "Sistema de salud chileno pasado, presente e incertidumbres del futuro," *Revista Chilena de Enfermedades Respiratorias*, vol. 39, no. 3, pp. 196–202, 2023.

[21] J. Eisenstein, *Introduction to natural language processing*. MIT press, 2019.

[22] R. Egger and J. Yu, "A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts," *Frontiers in sociology*, vol. 7, p. 886498, 2022.

[23] X. Liu, Q. Ma, J. Li, Z. Huang, X. Tong, T. Wang, H. Qin, W. Sui, and J. Luo, "Investigation of exercise interventions on postoperative recovery in lung cancer patients: A qualitative study using web crawling technology," *Patient Preference and Adherence*, pp. 1965–1977, 2024.

[24] R. Lawson, *Web scraping with Python*. Packt Publishing Ltd, 2015.

Anexo A. Resultados de modelos de aprendizaje supervisado con 3 clases

A continuación se muestran los resultados obtenidos para los modelos de aprendizaje supervisado empleando 3 clases (negativo, positivo y neutro).

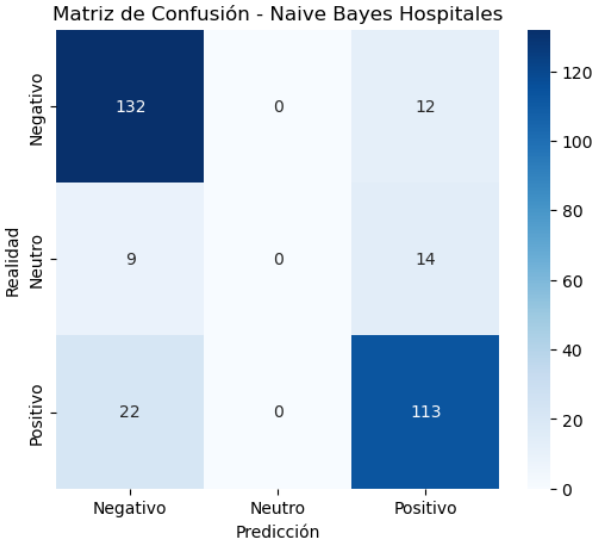


Figura A.28: Matriz de Confusión NB hospitales considerando 3 clases.

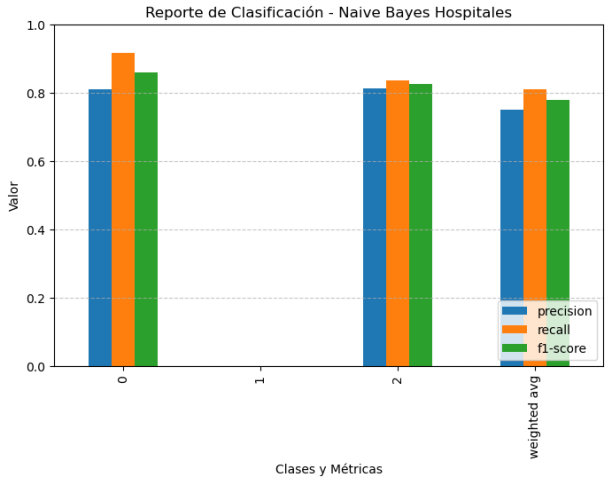


Figura A.29: Reporte de Clasificación NB hospitales considerando 3 clases.

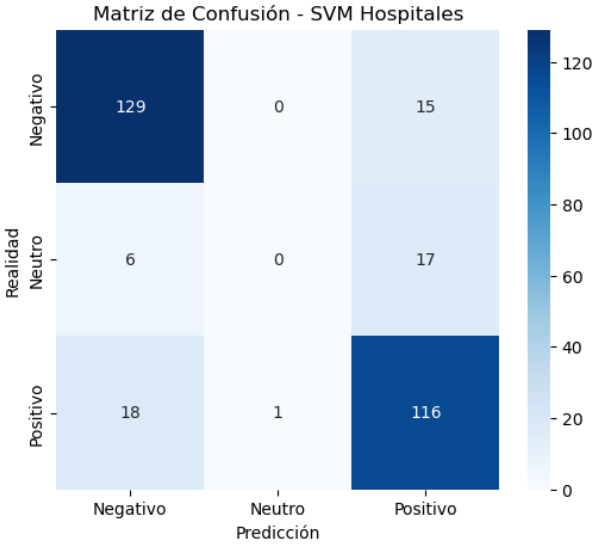


Figura A.30: Matriz de Confusión SVM hospitales considerando 3 clases.

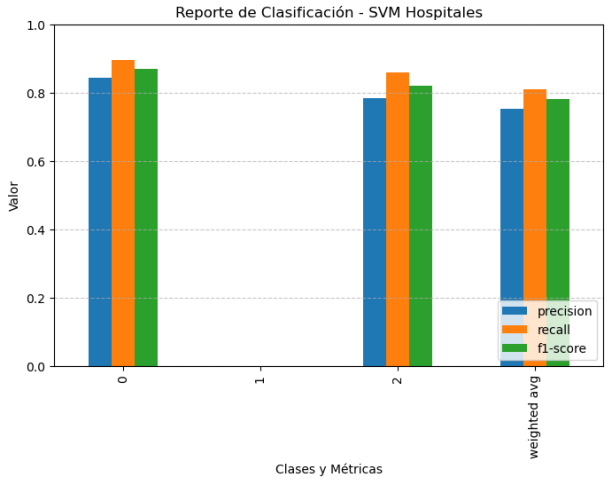


Figura A.31: Reporte de Clasificación SVM hospitales considerando 3 clases.

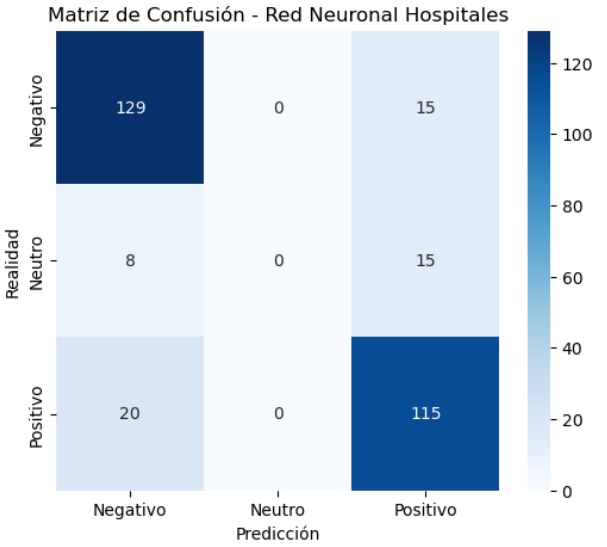


Figura A.32: Matriz de Confusión Red Neuronal hospitales considerando 3 clases.

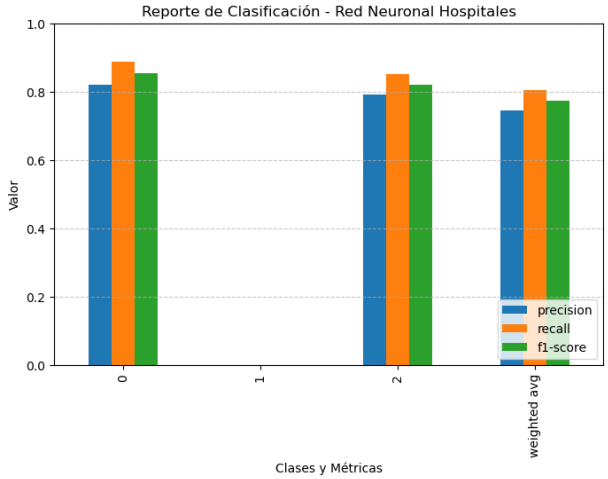


Figura A.33: Reporte de Clasificación Red Neuronal hospitales considerando 3 clases.

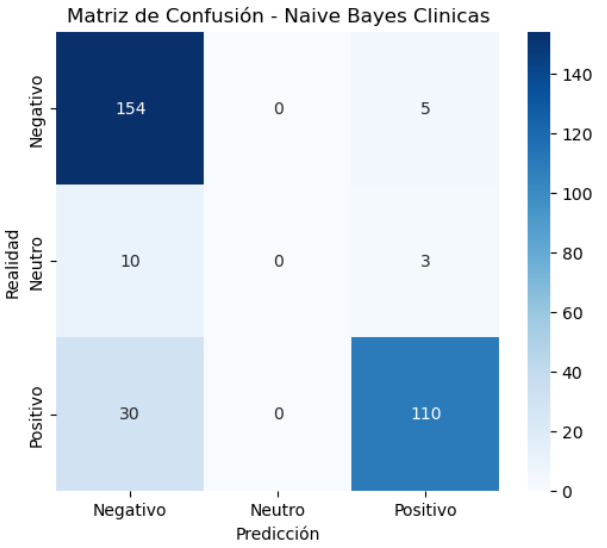


Figura A.34: Matriz de Confusión NB clinicas considerando 3 clases.

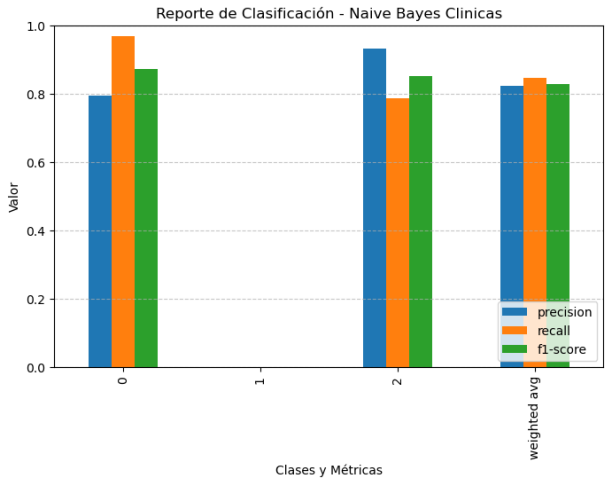


Figura A.35: Reporte de Clasificación NB clinicas considerando 3 clases.

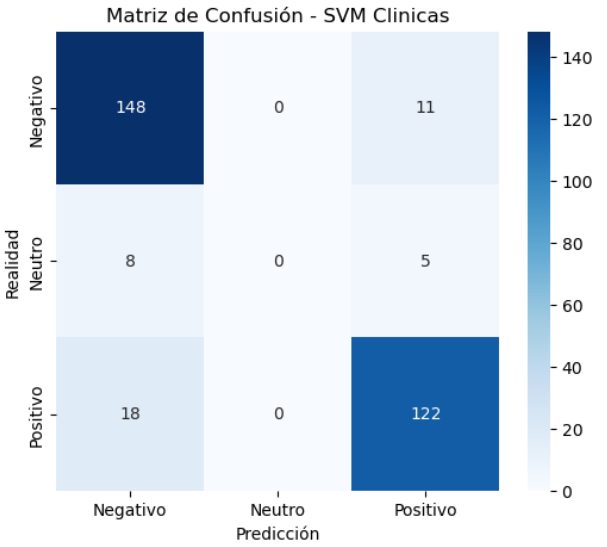


Figura A.36: Matriz de Confusión SVM clinicas considerando 3 clases.

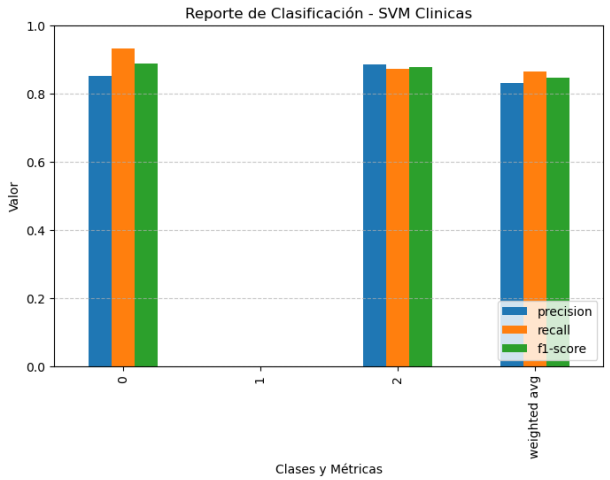


Figura A.37: Reporte de Clasificación SVM clinicas considerando 3 clases.

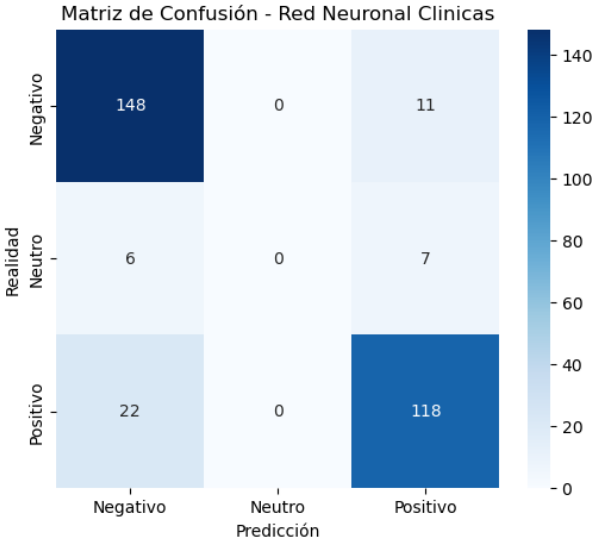


Figura A.38: Matriz de Confusión Red Neuronal clinicas considerando 3 clases.

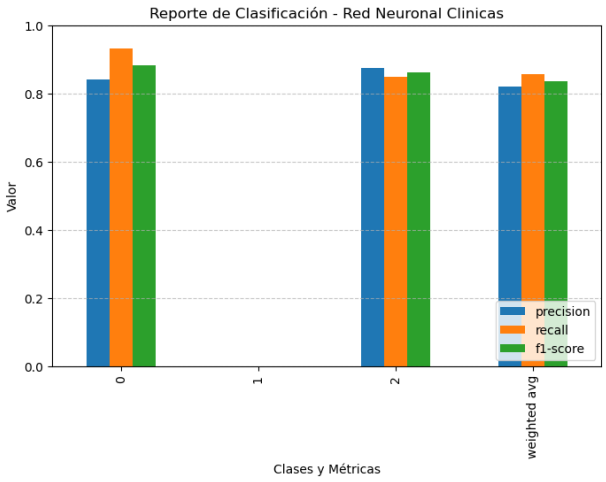


Figura A.39: Reporte de Clasificación Red Neuronal clinicas considerando 3 clases.

UNIVERSIDAD DE CONCEPCIÓN – FACULTAD DE INGENIERÍA
RESUMEN DE MEMORIA DE TÍTULO

Departamento	:	Departamento de Ingeniería Eléctrica
Carrera	:	Ingeniería Civil Biomédica
Nombre del memorista	:	Luis Sebastián Saavedra Acuña
Título de la memoria	:	Percepción de la calidad del servicio en el sistema de salud mediante análisis de opinión disponibles en la web.
Fecha de la presentación oral	:	9 de junio de 2025.
Profesor(es) Guía	:	Rosa Liliana Figueroa Iturrieta.
Profesor(es) Revisor(es)	:	Jean Paul Navarrete, Mario Medina.
Concepto	:	
Calificación	:	

Resumen

El presente trabajo tuvo como objetivo analizar la percepción de la calidad del servicio de salud en Chile a partir de opiniones de usuarios recopiladas en plataformas públicas en línea, utilizando técnicas de Procesamiento de Lenguaje Natural (PLN) y aprendizaje automático. La motivación principal radica en la falta de métodos automatizados de análisis de opinión en el contexto chileno, en contraste con la creciente disponibilidad de comentarios públicos en plataformas digitales. A través de web crawling automatizado, se recolectaron más de 2.000 opiniones de seis hospitales públicos y seis clínicas privadas, seleccionadas en base a criterios de tamaño poblacional y distribución geográfica.

Posteriormente, los textos fueron sometidos a un exhaustivo proceso de preprocesamiento, incluyendo la eliminación de acentos, normalización, tokenización y vectorización, para garantizar su correcta interpretación por los modelos analíticos. Para la clasificación de sentimientos, se implementaron tres algoritmos supervisados: Naive Bayes, Máquina de Vectores de Soporte (SVM) y una Red Neuronal Simple, utilizando la calificación numérica adjunta a cada opinión como criterio de etiquetado. Se decidió eliminar la clase neutra debido a su escasa representación, transformando el problema en una clasificación binaria entre sentimientos positivos y negativos. Los resultados mostraron que Naive Bayes obtuvo el mejor desempeño en hospitales, mientras que SVM lideró en clínicas, alineándose en parte con lo reportado en la literatura especializada.

En la segunda etapa, se aplicaron métodos de modelado de tópicos mediante aprendizaje no supervisado, utilizando Latent Dirichlet Allocation (LDA) y Factorización de Matrices No Negativas (FMN). Ambos enfoques permitieron identificar patrones temáticos recurrentes en las opiniones, tales como tiempos de espera, trato del personal y calidad de la atención, siendo FMN el que logró una mejor definición y segmentación de los tópicos, especialmente en los subconjuntos de opiniones negativas.

En conjunto, los resultados obtenidos evidencian que el uso de técnicas de PLN y machine learning puede automatizar de manera efectiva el análisis de la percepción ciudadana sobre el sistema de salud, proporcionando información valiosa para la mejora continua del servicio. Asimismo, el algoritmo desarrollado presenta potencial para ser implementado en sistemas de monitoreo en tiempo real, permitiendo a las instituciones de salud detectar tendencias en la opinión de los usuarios de manera temprana y oportuna, favoreciendo la toma de decisiones basada en evidencia.