
INSTITUTO TECNOLÓGICO SUPERIOR DE TEZIUTLÁN

Tesis



Desarrollo de un enfoque de análisis de medios sociales basado en técnicas de Procesamiento de Lenguaje Natural y análisis Psicolingüístico para determinar los principales factores de rechazo entre los movimientos antivacunas en Idioma Español

PRESENTA:

María Guadalupe Torrijos Rivera

CON NÚMERO DE CONTROL

21TE0006P

PARA OBTENER EL GRADO ACADÉMICO DE:

MAESTRA EN SISTEMAS COMPUTACIONALES

CLAVE DEL PROGRAMA ACADÉMICO

MPSCO-0127

DIRECTOR (A) DE TESIS:

María del Pilar Salas Zárate

CO- DIRECTOR DE TESIS:

“La Juventud de hoy, Tecnología del Mañana”

TEZIUTLÁN, PUEBLA, **JULIO 2023**

AGRADECIMIENTOS

Quiero expresar mi profundo agradecimiento a las siguientes personas e instituciones por su apoyo y contribución en la realización de esta tesis:

A Dios, por su constante guía y bendiciones en cada paso de este camino.

A mis hijos, Miguel y Aarón, por ser mi mayor fuente de inspiración y motivación. Su amor incondicional y comprensión fueron mi impulso para alcanzar este logro.

A mi esposo, Armando Barrientos, por su apoyo inquebrantable, paciencia y comprensión durante todo este proceso. Su amor y apoyo fueron fundamentales para superar los desafíos y obstáculos.

A mis padres, Gregoria Rivera y Armando Torrijos, por su amor, dedicación y sacrificio para brindarme una educación de mucho esfuerzo y sacrificio. Su apoyo incondicional y palabras de aliento fueron mi mayor fortaleza. A lo largo de mi vida han sido pieza clave para ser lo que hoy soy.

A mis hermanos, Mayra y Jesús Armando, por su apoyo y aliento constante. Su apoyo emocional y confianza en mis capacidades fueron fundamentales para alcanzar este logro.

A mis suegros, Gudelia y Artemio, por su cariño y respaldo. Desde que los conozco siempre me han mostrado su gran cualidad y grandeza de seres humanos. Siempre están dispuestos a brindarme su ayuda.

A mis amigas, Isabel, Magdalena, Martha y Paola, por estar siempre presentes, escucharme, brindarme su apoyo y aliento. Su amistad y compañía fueron un gran sostén en momentos difíciles.

Agradezco especialmente a mi asesora de tesis, María del Pilar Salas, por su valiosa orientación, conocimientos y paciencia. Su guía experta y retroalimentación fueron fundamentales para el desarrollo de esta investigación.

Agradezco a todos los profesores que tuve en la maestría, por su dedicación, enseñanzas y por compartir sus conocimientos y experiencias. Su compromiso académico fue fundamental para mi formación.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo financiero brindado a través de la beca de investigación. Su respaldo permitió llevar a cabo este estudio y contribuyó significativamente a su realización.

A todas estas personas e instituciones, mi más sincero agradecimiento por su apoyo, confianza y creencia en mí. Su contribución fue fundamental en el logro de esta meta.

Muchas gracias.

DEDICATORIA PERSONAL

A mis amados hijos, Miguel y Aarón,

Esta tesis está dedicada a ustedes, mis más grandes tesoros y fuente de inspiración. Cada paso que he dado en este camino ha sido motivado por ustedes. Sus sonrisas, amor incondicional y apoyo han sido mi motor para superar los desafíos y alcanzar mis metas. Quiero que sepan que su presencia en mi vida me llena de alegría y orgullo. Que esta tesis sea un ejemplo de que con determinación y esfuerzo se pueden lograr grandes cosas. Siempre estaré aquí para apoyarlos y guiarlos en cada uno de sus sueños.

A mi compañero de vida, Armando,

Quiero expresar mi más profundo agradecimiento por tu amor, apoyo incondicional y comprensión a lo largo de esta travesía. Tu constante aliento y paciencia han sido fundamentales en cada etapa de este proyecto. Gracias por ser mi roca, mi compañero y mi mayor admirador. Tu presencia a mi lado me ha dado la fuerza para superar los momentos difíciles y celebrar los logros. Esta tesis es un testimonio de nuestro gran trabajo en equipo. Agradezco infinitamente por estar a mi lado y compartir este viaje juntos.

A mis amados hijos y a mi esposo, esta tesis es un tributo a nuestra unidad como familia y al amor que nos une. Son mi mayor motivación y razón de ser. Su apoyo y comprensión han sido fundamentales en este camino. Que siempre recuerden que pueden lograr cualquier cosa que se propongan y que estoy aquí para apoyarlos en cada uno de sus sueños.

Con todo mi amor y gratitud.

ÍNDICE GENERAL

RESUMEN	12
ABSTRACT	14
INTRODUCCIÓN.....	16
CAPÍTULO I GENERALIDADES DEL PROYECTO.....	20
1.1 Marco teórico.....	20
1.1.1 Medios Sociales y Redes Sociales	20
1.1.2 Tipos de Redes Sociales	22
1.1.1.1 Redes sociales horizontales	22
1.1.1.2 Redes sociales verticales	22
1.1.1.3 Redes sociales de mensajería	23
1.1.3 Procesamiento del Lenguaje Natural	23
Aprendizaje supervisado	25
Aprendizaje Reforzado	26
1.1.4 Minería de Datos (Data Mining)	26
1.1.5 Modelos en Minería de Datos	29
1.1.6 Movimientos Antivacunas	31
1.1.7 Planteamiento del Problema	32
1.1.8 Justificación.....	33
1.1.9 Hipótesis	34
1.1.10 Objetivo general	35
1.1.11 Objetivos específicos	35
1.1.12 Alcances y limitaciones	36
CAPÍTULO II ESTADO DEL ARTE	37

2.1.1	Trabajos relacionados.....	37
2.1.2	Análisis comparativo de los trabajos relacionados.....	43
2.1.3	Propuesta de solución	44
4.1.3.1	Análisis de las Tecnologías de Información	44
4.1.3.2	Tecnologías Seleccionadas.....	45
	Python.....	45
	LIWC.....	45
	WEKA.....	46
CAPÍTULO III METODOLOGÍA Y DESARROLLO		47
3.1.1	Metodología de investigación/desarrollo	47
3.1.2	Comprensión del negocio.....	48
3.1.3	Comprensión de los datos.....	49
4.1.3.1	Exploración y análisis preliminar de los datos.....	50
4.1.3.2	Análisis de contenido.....	51
3.1.4	Preparación de los datos.....	52
4.1.4.1	Recopilación de datos.....	53
4.1.4.2	Limpieza de datos	54
4.1.4.3	Normalización de datos	55
4.1.4.4	Análisis Psicolingüístico.....	56
3.1.5	Modelado	62
3.1.6	Evaluación	67
3.1.7	Despliegue.....	70
CAPÍTULO IV RESULTADOS		70

4.1	Resultados obtenidos en YouTube.....	72
4.2	Resultados obtenidos en Facebook.....	75
4.3	Resultados obtenidos en Twitter	77
4.4	Análisis y Discusión de los Resultados.....	79
4.4.1	Pronombres Personales	81
4.4.1.1	Análisis Psicolingüístico.....	82
4.4.1.2	Análisis Psicológico.....	82
4.4.1.3	Análisis Cognitivo	82
4.4.2	Negación	83
4.4.2.1	Análisis Psicolingüístico.....	83
4.4.2.2	Análisis Psicológico.....	83
4.4.2.3	Análisis Cognitivo	84
4.4.3	Verbos.....	84
4.4.3.1	Análisis Psicolingüístico.....	84
4.4.3.2	Análisis Psicológico.....	85
4.4.3.3	Análisis Cognitivo	85
4.4.4	Familia	85
4.4.4.1	Análisis Psicolingüístico.....	86
4.4.4.2	Análisis Psicológico.....	86
4.4.4.3	Análisis Cognitivo	87
4.4.5	Emociones Positivas	87
4.4.5.1	Análisis Psicolingüístico.....	87
4.4.5.2	Análisis Psicológico.....	88

4.4.5.3	Análisis Cognitivo	88
4.4.6	Emociones Negativas.....	89
4.4.6.1	Análisis Psicolingüístico.....	89
4.4.6.2	Análisis Psicológico.....	90
4.4.6.3	Análisis Cognitivo	90
4.4.7	Ansiedad	91
4.4.7.1	Análisis Psicolingüístico.....	91
4.4.7.2	Análisis Psicológico.....	91
4.4.7.3	Análisis Cognitivo	92
4.4.8	Enfado	92
4.4.8.1	Análisis Psicolingüístico.....	93
4.4.8.2	Análisis Psicológico.....	93
4.4.8.3	Análisis Cognitivo	94
4.4.9	Cuerpo	94
4.4.9.1	Análisis Psicolingüístico.....	94
4.4.9.2	Análisis Psicológico.....	95
4.4.9.3	Análisis Cognitivo	95
4.4.10	Salud	96
4.4.10.1	Análisis Psicolingüístico.....	96
4.4.10.2	Análisis Psicológico.....	96
4.4.10.3	Análisis Cognitivo	97
4.4.11	Muerte	97
4.4.11.1	Análisis Psicolingüístico.....	98

4.4.11.2	Análisis Psicológico.....	98
4.4.11.3	Análisis Cognitivo	99
4.4.12	Signos de Puntuación, Interrogación y Exclamación	99
4.4.12.1	Análisis Psicolingüístico.....	99
4.4.12.2	Análisis Psicológico.....	100
4.4.12.3	Análisis Cognitivo	100
CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES.....		103
CAPÍTULO VI		105
PRODUCTOS ACADÉMICOS.....		105
REFERENCIAS		106

ÍNDICE DE FIGURAS

Figura 1.1 – Diferencia Medios y Redes Sociales y sus derivaciones **iError! Marcador no definido.**

Figura 1.2 – Etapas de Análisis en el PLN..... **iError! Marcador no definido.**

Figura 1.3 – Tipos de Aprendizaje Automático..... **iError! Marcador no definido.**

Figura 1.5 Flujo del Proceso de la Minería de Datos**iError! Marcador no definido.**

Figura 1.6 Modelos que ocupa la Minería de Datos **iError! Marcador no definido.**

ÍNDICE DE TABLAS

Tabla 1. Ejemplo de tabla. **iError! Marcador no definido.**

RESUMEN

Esta investigación presenta el desarrollo de un enfoque integral para analizar datos de medios sociales y examinar los principales factores que contribuyen al rechazo de la vacunación dentro de los movimientos antivacunas de habla hispana. El estudio emplea un marco multidisciplinario que combina técnicas de Procesamiento de Lenguaje Natural (PLN) y Psicolingüística, siguiendo la metodología CRISP (Cross-Industry Standard Process for Data Mining).

Los objetivos de la investigación abarcan varias etapas. En primer lugar, se realiza una amplia revisión del estado del arte para identificar las herramientas de software más utilizadas y las técnicas relevantes en PLN y Psicolingüística para analizar opiniones y emociones en textos en español. Esto sirve como base para el análisis posterior.

Utilizando la metodología CRISP, se identifican y recopilan fuentes relevantes de datos de medios sociales, como Twitter, Facebook y YouTube de los cuales se obtuvieron los comentarios más relevantes respecto a los movimientos antivacunas. Estas fuentes se utilizan luego para construir un corpus de opiniones y discursos sobre la vacunación, proporcionando un conjunto de datos rico para la investigación.

La investigación continúa con la actividad de identificar y categorizar automáticamente las palabras en un texto de acuerdo con diversas dimensiones psicológicas y lingüísticas. Con esta técnica de PLN se realiza la extracción de características psicolingüísticas para analizar el contenido de textos en diferentes dimensiones lingüísticamente significativas.

Después de la etapa de preparación de los datos, se selecciona el algoritmo de aprendizaje automático más adecuado en función de la naturaleza de los datos y los objetivos de la investigación. Este algoritmo se entrena utilizando el corpus desarrollado para crear modelos de predicción destinados a identificar los principales factores que contribuyen al rechazo de la vacunación en los movimientos antivacunas de habla hispana.

Aplicando la metodología CRISP, los modelos de predicción resultantes se evalúan y analizan minuciosamente para extraer ideas valiosas sobre los factores subyacentes que impulsan el rechazo de las vacunas dentro del contexto hispanohablante. Estos hallazgos tienen implicaciones significativas para comprender las motivaciones y creencias dentro de los movimientos antivacunas, y pueden informar el diseño de estrategias de comunicación y educación dirigidas para promover una mayor aceptación de las vacunas.

En conclusión, esta tesis presenta un enfoque innovador que combina PLN y Psicolingüística dentro del marco de la metodología CRISP para analizar datos de medios sociales como Twitter, Facebook y YouTube en el contexto de los movimientos antivacunas de habla hispana. La aplicación de este enfoque contribuye a una comprensión más profunda de los factores que influyen en el rechazo de las vacunas y facilita la toma de decisiones basada en evidencia en las intervenciones de salud pública.

ABSTRACT

This research presents the development of a comprehensive approach to analyze social media data and examine the key factors contributing to vaccine rejection within Spanish-speaking anti-vaccine movements. The study employs a multidisciplinary framework that combines Natural Language Processing (NLP) and Psycholinguistics techniques, following the CRISP (Cross-Industry Standard Process for Data Mining) methodology.

The research objectives span several stages. Firstly, an extensive state-of-the-art review is conducted to identify the most widely used software tools and relevant techniques in NLP and Psycholinguistics for analyzing opinions and emotions in Spanish texts. This serves as the foundation for the subsequent analysis.

Using the CRISP methodology, relevant sources of social media data, such as Twitter, Facebook, and YouTube are identified and collected, from which the most relevant comments regarding anti-vaccine movements are obtained. These sources are then used to construct a corpus of opinions and discourses about vaccination, providing a rich dataset for research purposes.

The research proceeds with the task of automatically identifying and categorizing words in a text according to various psychological and linguistic dimensions. This NLP technique enables the extraction of psycholinguistic features to analyze the content of texts in different linguistically significant dimensions.

Following the data preparation stage, the most suitable machine learning algorithm is selected based on the nature of the data and the research objectives. This algorithm is trained using the developed corpus to create prediction models aimed at identifying the key factors contributing to vaccine rejection within Spanish-speaking anti-vaccine movements.

Applying the CRISP methodology, the resulting prediction models are thoroughly evaluated and analyzed to extract valuable insights into the underlying factors driving vaccine rejection within the Spanish-speaking context. These findings have

significant implications for understanding motivations and beliefs within anti-vaccine movements and can inform the design of targeted communication and education strategies to promote greater vaccine acceptance.

In conclusion, this thesis presents an innovative approach that combines NLP and Psycholinguistics within the framework of the CRISP methodology to analyze social media data from platforms like Twitter, Facebook, and YouTube within the context of Spanish-speaking anti-vaccine movements. The application of this approach contributes to a deeper understanding of the factors influencing vaccine rejection and facilitates evidence-based decision-making in public health interventions.

INTRODUCCIÓN

Los movimientos antivacunas tienen ya una larga trayectoria en el transcurso de la historia, las raíces se cimentaron hace más de un siglo. Podemos decir que no es algo nuevo y que la negación y la desconfianza han aumentado y se ha vuelto un fenómeno mundial. En el siglo XIX en el Reino Unido surgen los primeros grupos organizados antivacunación; su progreso ha tenido un desarrollo evolutivo y siguen vigentes.

Haciendo un poco de remembranza de lo que ha acontecido a nivel internacional por la Pandemia de COVID 19 han sucedido diferentes cambios sociales y emocionales; se puede decir que hasta culturales. Hay una gran modificación en nuestra forma de relacionarnos, de entablar comunicación y hasta la forma de trabajar ha marcado una evolución empresarial. Quizá muchos de estos cambios se mantengan durante mucho tiempo y algunos sean adaptables para mejorar técnicas de vida tecnológica y social. Entre todos estos cambios también se puede decir que la confianza en las vacunas y su innovación reactiva han provocado un gran revuelo en la sociedad. Pero no todas las personas confiaron en esta aplicación masiva de inoculación y la batalla del virus, muchas de ellas se declararon "antivacunas" y se manifestaron a través de los medios sociales para dar peso a sus convicciones.

La presente investigación se encamina en desarrollar un enfoque de análisis de medios sociales basado en técnicas de Aprendizaje automático y análisis psicolingüístico para determinar los principales factores de rechazo de los movimientos antivacunas.

Los medios sociales son todos los instrumentos, herramientas, plataformas, aplicaciones, tecnología para crear conversación y comunicarnos. (Sergio Galán, 2016). Estos serán utilizados como instrumentos de estudio para recopilar (extraer) las opiniones, sentimientos, manifestaciones e ideas que permitan realizar un análisis de las características psicolingüísticas a través de técnicas de procesamiento

de lenguaje y obtener conocimiento para determinar los elementos de resistencia en los movimientos antivacunas.

El presente documento se encuentra estructurado de la siguiente manera; **en el capítulo 1**, se abordan los antecedentes del proyecto en cual se realiza una síntesis conceptual de los trabajos realizados previamente que sirven de base para la investigación de esta tesis.

Además, se aborda el marco teórico y todos los conceptos que han servido de apoyo para fortalecer y dar cimiento a la estructura de contenido del proyecto. Se presenta el planteamiento del problema, el cual es la base de este esquema de investigación y determina de forma clara y concreta el propósito de dar respuesta al estudio presentado. También se indica el objetivo general y los objetivos específicos los cuales nos llevan a cumplir la meta principal y los pasos que se deben dar para poder cumplir dicha meta. Se plantea la justificación en la cual se explica la razón por la cual fue elegido el tema de investigación y la aportación que se quiere lograr con el enfoque utilizado.

En el capítulo 2, se realiza una descripción general, un resumen y una evaluación (análisis comparativo) del estado actual del conocimiento sobre los trabajos relacionados que ya existen y que han servido de fundamento para comparar métodos paralelos con el propósito de referir varios puntos y comprender el problema de estudio para conseguir alternativas de solución que giran en torno a la investigación. Se realiza de forma gráfica el análisis de comparación lo cual simplifica la estrategia de estudio entre las diversas investigaciones presentadas y que muestran los resultados obtenidos en cada una de ellas y ayudan a observar las semejanzas o diferencias que existen entre las mismas y sirven para el objeto de estudio.

En el capítulo 3, en este capítulo, se muestra el desarrollo de la metodología empleada en la presente investigación. La metodología constituye el pilar fundamental de cualquier estudio científico. El objetivo principal es describir en detalle el diseño metodológico adoptado para llevar a cabo el estudio, se expondrá

la estructura del diseño de investigación, la selección de la muestra, la recopilación y análisis de datos, así como los procedimientos utilizados para garantizar la validez y la confiabilidad de los resultados.

Se estructura en varias secciones, cada una de ellas enfocada en un aspecto específico de la metodología. Se inicia con una descripción del diseño general de la investigación, seguida por la presentación de los procesos utilizados para la recolección de datos. Asimismo, se detallarán las actividades para la selección de la muestra y se explicará la estrategia de análisis de datos adoptada.

En el capítulo 4, constituye el núcleo central, donde se exponen y analizan detalladamente los resultados obtenidos a partir de la metodología previamente descrita. El objetivo principal es brindar una síntesis clara y precisa de los resultados alcanzados, basados en un análisis riguroso y fundamentado en la evidencia recopilada. Se describirán los datos relevantes, se presentarán los resultados de acuerdo al análisis obtenido y se discutirán los hallazgos a la luz de los objetivos propuestos. Es un elemento esencial de la tesis, en el que se exponen los frutos de la investigación y se brindan conclusiones significativas para el progreso del conocimiento.

En el capítulo 5, es donde se marca el punto culminante de la investigación ya que se presentan las conclusiones obtenidas a partir del análisis riguroso de los resultados y se ofrecen recomendaciones fundamentadas en los hallazgos del estudio. Representa el cierre y se proporciona una síntesis clara y concisa de los resultados, destacando los hallazgos más relevantes y su contribución al conocimiento en el campo de estudio. Se realizará una reflexión sobre cómo los resultados obtenidos se relacionan. Este capítulo constituye un cierre sólido y fundamentado para la presente tesis, enriqueciendo el corpus de conocimiento en el campo de estudio.

En el capítulo 6, representa una oportunidad para mostrar la trascendencia de la investigación más allá de la tesis en sí misma, al presentar los productos académicos derivados del trabajo investigativo. Los productos académicos son una vía para

compartir el conocimiento generado y contribuir al avance de la disciplina, enriqueciendo el panorama académico y científico en el campo de estudio.

CAPÍTULO I GENERALIDADES DEL PROYECTO

1.1 Marco teórico

Estos últimos años han marcado una línea muy clara entre la prevención para atacar una enfermedad y la de poner freno a una pandemia declarada a nivel mundial. Pero, ¿qué tan preparada esta la sociedad para aceptar o no una vacuna como solución ante esta situación que parecía salir de control? Los medios sociales han sido una plataforma que desahoga gran cantidad de contenido con comentarios y opiniones en las cuales de forma libre las personas pueden expresar los motivos por los cuales no existe confianza en las vacunas; y esto dio nuevo peso a los movimientos antivacunas que ya existen desde años atrás.

Existen varias investigaciones que se han realizado con base en las plataformas de las redes sociales donde se libra una batalla entre aquellos que sustentan bajo sus argumentos sobre lo que beneficia una vacuna y también lo que se opina de forma contraria o negativa.

En el presente apartado se exponen los principales conceptos que son utilizados en la elaboración de la tesis. Como primer paso es conocer e identificar qué son los medios sociales y determinar la importancia de su influencia en la audiencia pública que cada día es más participativa. El siguiente paso es comprender el termino Procesamiento del Lenguaje Natural (PLN) e incluir cada una de las ramas conceptuales que de ella derivan. Posterior a esto entender lo que es el análisis psicolingüístico y determinar las técnicas para extraer la información y los modelos predictivos para la obtención de resultados. Y finalmente conceptualizar lo que son los movimientos antivacunas y cómo impactan a través de los medios sociales.

1.1.1 Medios Sociales y Redes Sociales

Siempre que hacemos referencia a los medios sociales, normalmente estamos hablando de las redes sociales. Sin embargo, no tratan exactamente de lo mismo.

Podemos decir que los “medios” son las herramientas y “redes” es el resultado de lo que se genera o crea a través de ellas.

Por lo tanto, para poder conceptualizar y distinguir estos dos significados se concluye que los medios sociales son la plataforma utilizada por comunidades de usuarios en la cual se puede crear contenido, compartirlo o bien realizar conversaciones. Y las redes sociales son la asociación que existe cuando dos o más usuarios se relacionan a través de un medio social.

Actualmente las redes sociales son un portal de diversidad que a través de ellas la información se difunde de una manera tan abierta y tan acelerada que es difícil distinguir lo que es real y lo que es falso en su contenido. Esto puede generar un impacto positivo o negativo en las personas; por tal motivo es considerada como base para realizar diferentes estudios sobre los sentimientos que se manifiestan en los comentarios compartidos. Podemos resumir que los medios sociales son los conductos de comunicación y las redes sociales son los enlaces en los que se mantiene un diálogo entre comunidades (usuarios) con intereses comunes (Fernández Rivera, 2019). En la **Figura 1.1** se puede observar la diferencia entre medios y redes sociales, además de la derivación que existen en las mismas.

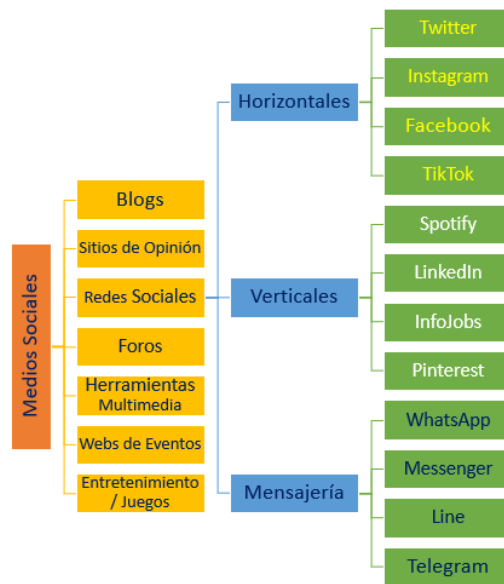


Figura 1.1 – Diferencia Medios y Redes Sociales y sus derivaciones Fuente: Propia

1.1.2 Tipos de Redes Sociales

No todas las redes sociales son iguales, aunque el objetivo de comunicarse sea el mismo, todo dependerá según la funcionalidad que se realiza en ellas. Por tal razón podemos clasificarlas de la siguiente manera:

1.1.1.1 Redes sociales horizontales

Estas plataformas son de participación libre y son reconocidas como comunidades generalistas. Una temática concreta no existe; y por esta razón su dirección es a un público sin un fin concreto. Algunos ejemplos característicos de estas redes son Facebook, Instagram, Pinterest, Twitter (NextVisiion, 2021).

1.1.1.2 Redes sociales verticales

La peculiaridad de estas redes es que las personas se agrupan (comunidades) en función a un tema determinado. Manejan un tipo de restricción; ya que puede ser solicitada una "invitación" para poder ingresar

a las mismas. Su objetivo primordial radica en obtener información de un tema en particular y ofrecer experiencia y conocimiento común con otros miembros de la comunidad (NextVisiion, 2021).

1.1.1.3 Redes sociales de mensajería

Las redes sociales de mensajería muchas veces son consideradas como redes sociales verticales. Sin embargo, el gran auge de éstas en los últimos años hace que pertenezcan a una nueva clase. La mensajería instantánea en un fenómeno que ha dado un nuevo giro y está transformando las redes sociales; ya que la manera en que las personas interactúan es de una forma muy dinámica y acelerada. Las más utilizadas y más comunes son: WhatsApp, Messenger, Line, Telegram, WeChat Discord, y varias más (NextVisiion, 2021).

1.1.3 Procesamiento del Lenguaje Natural

El Procesamiento del lenguaje Natural (PLN) pertenece a una de las ramas de la inteligencia artificial y la lingüística, y tiene como objetivo principal dotar a las computadoras con la capacidad de entender, interpretar y manipular el lenguaje humano (VIEWNEXT, 2020). Esta tecnología tan completa e innovadora ha alcanzado un alto nivel en el mundo de la informática gracias a la aplicación del aprendizaje automático. De forma más concreta, se enfoca en el procesamiento de las comunidades humanas, fragmentándola en partes y considerando los elementos más importantes del mensaje.

Podemos decir que utilizar PLN es una fórmula para ocupar los mecanismos computacionales en que las máquinas y el ser humano se pueden comunicar por medio de programas que simulen dicha comunicación o bien se obtengan las ventajas proporcionadas por el mismo (decide, 2019).

En la **Figura 1.2** se pueden observar las etapas (de análisis) que pueden ser ocupadas para el procesamiento del lenguaje natural. Sin embargo, puede suceder que no todas apliquen; todo dependerá del objetivo de su aplicación.

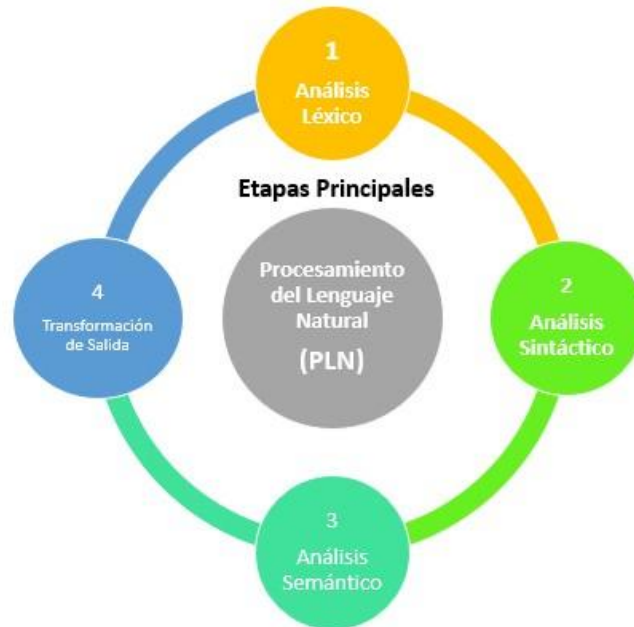


Figura 1.2 – Etapas de Análisis en el PLN Fuente: Propia

Análisis morfológico o léxico. Radica en el análisis interno de las palabras que crean oraciones para sustraer rasgos flexivos, marcas y unidades léxicas. Este tipo de análisis es fundamental para hacer consistente la información básica: categoría sintáctica y significado léxico.

Análisis sintáctico. Realiza un análisis de la estructura de oraciones en base al modelo gramatical utilizado (estadístico o lógico).

Análisis semántico. En este se genera la interpretación de las oraciones, y esto se logra eliminando las ambigüedades morfosintácticas.

Transformación de Salida (Análisis pragmático). En la interpretación final se incorpora el análisis del contexto de uso.

Aprendizaje Automático (Machine Learning)

En el campo de la Inteligencia Artificial existe una disciplina que, por medio de algoritmos, proporciona a las computadoras la capacidad de identificar patrones en grandes volúmenes de datos y generar predicciones (análisis predictivo). Con este aprendizaje se permite a las computadoras ejecutar tareas definidas en forma autónoma, es decir, no se necesita de una programación para ser elaboradas (IBERDROLA, 2022). En la **Figura 1.3** se presentan los diferentes tipos de aprendizaje automático que pueden ser utilizados dependiendo la tarea que se quiera realizar.



Figura 1.3 – Tipos de Aprendizaje Automático Fuente: Propia

Aprendizaje supervisado

Es un modelo en el cual el proceso para la generación de conocimiento se lleva a cabo considerando un grupo de ejemplos o datos etiquetados, y los resultados que proyecta la operación son previamente identificados. Este tipo de modelo aprende de los resultados arrojados y añade ajustes en los parámetros interiores para lograr la adaptación de datos nuevos que van ingresando al sistema (GRAPH Everywhere, 2022).

Aprendizaje No Supervisado

En este modelo son considerados conjuntos de datos sin etiquetar y; de los cuales no se tiene conocimiento previo de la estructura que tienen. El objetivo principal de este tipo de aprendizaje es obtener información relevante o importante sin conocimiento previo de la referencia en las variables de salida. Además se realiza exploración en la estructura de los datos que no están etiquetados (GRAPH Everywhere, 2022).

Aprendizaje Reforzado

Este tipo de aprendizaje tiene como finalidad elaborar modelos que acrecienten el rendimiento, considerando como base la respuesta que se genera en cada interacción ejecutada. El resultado obtenido es el producto de una operación correcta o conjunto de datos arrojados que forman parte de una medida específica. Este aprendizaje es considerado dentro de lo que se conoce como Deep Learning o aprendizaje profundo (GRAPH Everywhere, 2022).

1.1.4 Minería de Datos (Data Mining)

La minería de datos es un proceso que ayuda identificar la información que puede ser procesada en grandes conjuntos de datos. Hace uso del análisis matemático para suponer los patrones y tendencias de relación que existe entre los datos. Regularmente, estos patrones no son detectables por medio de una exploración de forma habitual ya que las relaciones resultan demasiado complicadas, o porque el volumen de datos es muy grande (Learn Microsoft, 2022). En la **Figura 1.4** Se puede observar el flujo de la Minería de Datos

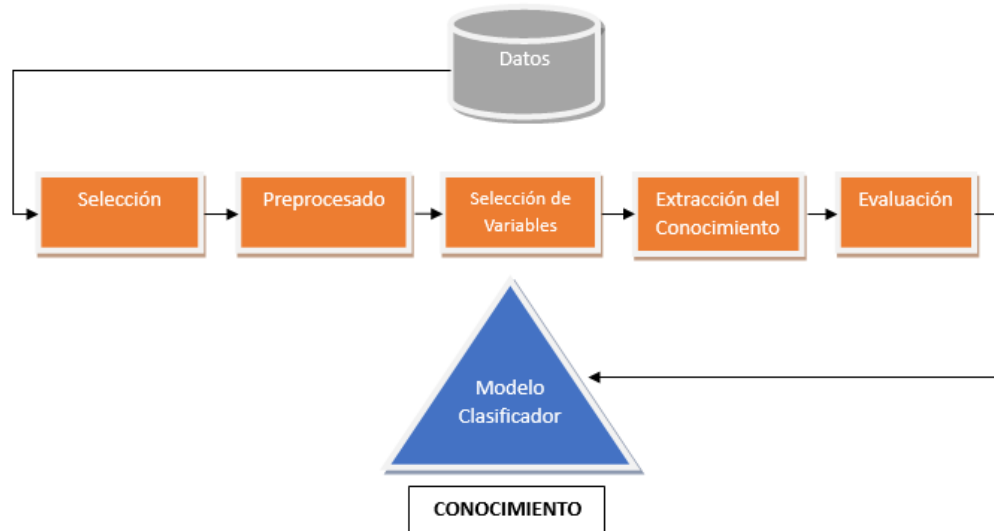


Figura 1.4 - Flujo del Proceso de la Minería de Datos Fuente: Propia

- **Datos:** Son el componente fundamental sobre el cual se realiza el análisis y se extrae conocimiento. Los datos representan información en bruto recopilada de diversas fuentes, como bases de datos, registros, archivos, sensores, redes sociales, entre otros.
- **Selección:** Se refiere al proceso de elegir una muestra o subconjunto específico de datos de un conjunto de datos más grande para realizar el análisis o modelado. La selección de datos se utiliza cuando el conjunto completo de datos es demasiado grande o no es necesario para el análisis o la tarea específica que se está realizando.
- **Preprocesamiento:** Es una etapa crucial que implica la preparación y transformación de los datos antes de aplicar técnicas de análisis o modelado. El objetivo principal del preprocesamiento es mejorar la calidad de los datos, eliminar ruido y asegurar que los datos sean adecuados para su análisis y extracción de conocimientos.

El preprocesamiento de datos puede involucrar una serie de pasos y técnicas, que incluyen:

- **Limpieza de datos:** Se refiere a la eliminación o corrección de datos incorrectos, inconsistentes o incompletos. Esto implica tratar valores faltantes, outliers (valores atípicos o valores extremos), duplicados o datos contradictorios que puedan afectar la precisión de los resultados.
- **Integración de datos:** En ocasiones, los datos pueden provenir de múltiples fuentes y es necesario combinarlos en un único conjunto de datos coherente. Esto puede implicar la resolución de problemas de formato, representación o esquema.

- **Transformación de datos:** Se refiere a la modificación de los datos para que se ajusten mejor a los requisitos del análisis o de los algoritmos de minería de datos. Esto puede incluir la normalización o estandarización de variables numéricas, la discretización de variables continuas, la codificación de variables categóricas o la transformación de datos textuales en representaciones numéricas.
- **Selección de características:** Como se mencionó anteriormente, la selección de características implica elegir las variables más relevantes para el análisis y el modelado. Esto puede implicar la eliminación de características redundantes, irrelevantes o de baja importancia para mejorar la eficiencia y la interpretación de los modelos.
- **Reducción de dimensionalidad:** En conjuntos de datos con un gran número de características, la reducción de dimensionalidad se utiliza para representar los datos en un espacio de menor dimensión. Esto puede ayudar a reducir la complejidad, el tiempo de procesamiento y mejorar el rendimiento del análisis y la modelización.
- **Selección:** Se refiere al proceso de elegir las características o variables más relevantes o significativas de un conjunto de datos para su análisis y modelado. Se busca identificar las variables que tienen mayor poder predictivo o que contribuyen de manera significativa a los patrones y relaciones presentes en los datos. Al seleccionar un subconjunto relevante de características, se pueden obtener modelos más simples, comprensibles y eficientes, evitando la redundancia y el ruido en los datos.
- **Extracción del conocimiento:** Se refiere al proceso de descubrir información valiosa, patrones, relaciones o conocimientos útiles a partir de un conjunto de datos. Este proceso implica aplicar técnicas de análisis y modelado a los datos para obtener conocimientos significativos y accionables que puedan utilizarse para tomar decisiones informadas o generar valor.
- **Evaluación:** Es el proceso de medir y analizar la calidad y el rendimiento de los modelos y resultados obtenidos durante el análisis de los datos. Esta etapa es fundamental para determinar la efectividad y utilidad de los modelos generados, así como para comprender la precisión y confiabilidad de los conocimientos extraídos.
- **Modelo Clasificador y generación de conocimiento:** Se utilizan algoritmos de aprendizaje automático para construir modelos que puedan predecir o clasificar eventos o resultados futuros. Estos modelos pueden revelar patrones o relaciones que no eran aparentes en los datos originales y permiten obtener conocimientos sobre el comportamiento y las influencias en el fenómeno estudiado.

1.1.5 Modelos en Minería de Datos

Los modelos predictivos y descriptivos en la minería de datos son dos enfoques diferentes para analizar los datos y extraer conocimientos. En la **Figura 1.5** se puede observar los modelos que ocupa la Minería de Datos.

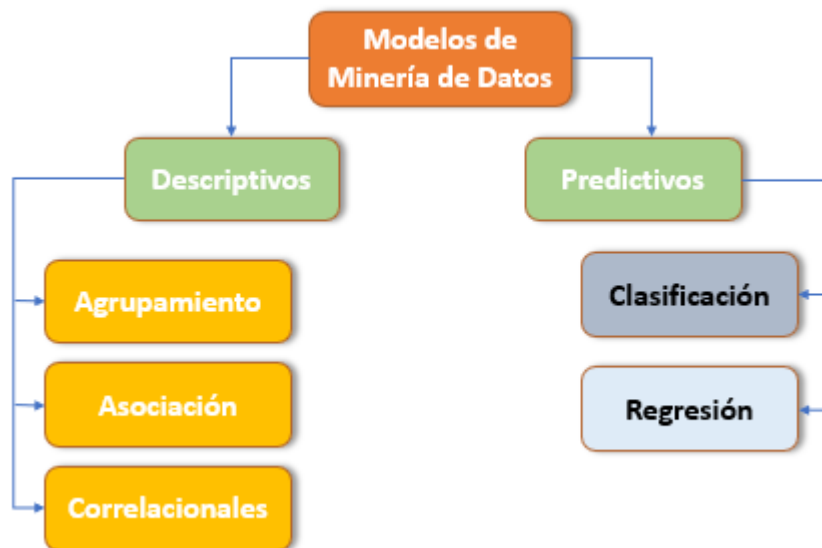


Figura 1.5 - Modelos que ocupa la Minería de Datos Fuente: Propia

Los **modelos descriptivos** se utilizan para resumir y describir los datos existentes sin hacer predicciones sobre nuevos datos. Estos modelos buscan identificar patrones, tendencias y relaciones dentro de los datos que proporcionen una comprensión más profunda del fenómeno estudiado (Learn Microsoft, 2022).

- **Agrupamiento:** es una herramienta valiosa para descubrir estructuras y relaciones en los datos. Proporciona una forma de organizar, resumir y comprender mejor los conjuntos de datos, lo que puede llevar a una toma de decisiones más informada y a la identificación de patrones valiosos que de otro modo podrían pasar desapercibidos.
- **Asociación:** Se utiliza para descubrir patrones y relaciones interesantes entre diferentes elementos o atributos en un conjunto de datos. Ayuda a comprender las asociaciones frecuentes y las co-ocurrencias, lo que puede llevar a una mejor comprensión del comportamiento de los usuarios, la toma

de decisiones más informadas y la generación de recomendaciones personalizadas.

- **Correlacionales:** Se utiliza para analizar la relación estadística entre variables. Ayuda a comprender la dependencia, seleccionar características relevantes, predecir valores faltantes y analizar tendencias en los datos. Sin embargo, es importante tener en cuenta que la correlación no implica causalidad directa, y se requiere un análisis adicional para establecer relaciones causales.

La creación de un **modelo predictivo** en la minería de datos se realiza por medio de la aplicación de un algoritmo de datos. Sin embargo, no solo es un algoritmo o un contenedor de metadatos, se puede decir que es un grupo de datos, en la cual se generan estadísticas y un determinado número de patrones que pueden ser aplicados a nuevos datos para generar predicciones y suponer relaciones (Learn Microsoft, 2022).

- **Clasificación:** Se utiliza para asignar categorías o etiquetas a nuevas instancias o ejemplos. Tiene una amplia gama de aplicaciones, incluyendo clasificación de correos electrónicos, detección de fraudes, diagnóstico médico, detección de intrusiones, clasificación de documentos y recomendación de productos. Ayuda a automatizar tareas de clasificación y toma de decisiones, mejorando la eficiencia y la precisión en diferentes dominios.
- **Regresión:** Se utiliza para predecir valores numéricos o continuos en función de variables independientes o predictores. En lugar de asignar categorías o clases como en el modelo de clasificación, el modelo de regresión busca establecer una relación matemática entre las variables independientes y la variable dependiente, que es el valor que se desea predecir. El objetivo del modelo de regresión es encontrar una función o ecuación que pueda describir y predecir la relación entre las variables. A través del análisis de los datos disponibles, el modelo de regresión estima los coeficientes o parámetros que mejor se ajustan a los datos y luego utiliza esta función para realizar predicciones sobre nuevos datos.

Estos son solo algunos ejemplos de modelos utilizados en minería de datos. Cada modelo tiene sus propias ventajas y limitaciones, y la elección del modelo depende del tipo de problema, los datos disponibles y los objetivos de predicción que se buscan y persiguen.

1.1.6 Movimientos Antivacunas

Estos movimientos pueden ser definidos como el grupo de individuos que por diferentes motivos consideran y creen que las vacunas y su aplicación afectan al ser humano; y por consecuencia generan un mayor perjuicio en la salud que el posible beneficio que se pudiera obtener de las mismas. Estos grupos colectivos se mantienen activos y comparten información sin fundamento científico; pero de manera convincente entre personas con su misma ideología (López Santamaría, 2015).

Los argumentos y creencias de los movimientos antivacunas varían, pero suelen incluir afirmaciones de que las vacunas son ineficaces, causan efectos secundarios graves o están relacionadas con enfermedades crónicas. Algunos movimientos antivacunas también se oponen a las políticas de vacunación obligatoria y defienden el derecho individual de decidir si vacunarse o no.

Los movimientos antivacunas pueden tener diferentes motivaciones y creencias. Algunas personas pueden estar preocupadas por los efectos secundarios de las vacunas y creen que representan un riesgo mayor que la enfermedad misma. Otros pueden tener creencias religiosas o filosóficas que se oponen a la intervención médica, incluso en forma de vacunas. También existen teorías de conspiración que afirman que las vacunas son parte de una agenda oculta o que están diseñadas para causar daño a la población.

A pesar de que la inmunización es una de las intervenciones más exitosas y seguras en la historia de la salud pública, los movimientos antivacunas han ganado atención y seguidores en los últimos años, en parte debido a la difusión de información a través de las redes sociales y la facilidad de compartir opiniones y testimonios personales en línea.

La propagación de información errónea y desinformación por parte de los movimientos antivacunas puede tener consecuencias negativas para la salud pública.

1.1.7 Planteamiento del Problema

Las redes sociales han evolucionado de una forma tan impactante que se han convertido en la estructura perfecta de comunicación en un mundo digital. Son una herramienta que usamos día con día y se tiene acceso prácticamente desde cualquier lugar del mundo. Sin embargo, influyen tanto y de una forma significativa en la vida cotidiana de las personas que pueden ser un factor positivo o bien un escenario no favorable para la sociedad. El mal uso de estos puede afectar en la veracidad de la información y desencadenar una dispersión de mitos o fundamentos establecidos solo desde la perspectiva personal de uno o varios individuos con el mismo pensamiento.

En la actualidad el uso de las vacunas forma una de las medidas sanitarias con mayor beneficio para el ser humano. Pero también existe la contraparte que desacredita y no está de acuerdo en las ventajas que puede aportar la aplicación de las mismas. Estos movimientos antivacunas difunden información convincente y casi siempre sin bases científicas que siembran desconfianza en la población. Por consecuencia a lo antes planteado se pretende realizar un estudio de los comentarios que tienen relación con las vacunas en las redes sociales y se pueda ayudar a identificar todos aquellos factores de rechazo que contribuyen a la desconfianza en las vacunas. Por tal razón se busca dar a conocer el conocimiento obtenido al sector salud y a la sociedad misma; y puedan ocupar la información para implementar prácticas de comunicación y se pueda reducir el impacto de datos falsos.

Existen varias investigaciones que han abordado el análisis de sentimientos sobre las opiniones en pro y en contra de las vacunas en diversas partes de mundo. Muchos de estos estudios fueron realizados en Europa (en su mayoría) y hasta el momento en América Latina son pocos los dedicados a ello. Por consecuencia,

el presente estudio considera como insumo los comentarios en idioma español de las redes sociales en países latinos.

1.1.8 Justificación

La propagación de movimientos antivacunas en las redes sociales ha generado preocupación en la comunidad científica y las autoridades sanitarias. Estos movimientos desafían la efectividad de las vacunas y promueven información errónea, lo que puede tener graves consecuencias para la salud pública. Es fundamental comprender los factores que contribuyen al rechazo de las vacunas y desarrollar estrategias efectivas para abordar este problema.

El presente estudio justifica la necesidad de desarrollar un enfoque de análisis de medios sociales basado en técnicas de Aprendizaje automático y análisis Psicolingüístico para determinar los principales factores de rechazo de los movimientos antivacunas en opiniones y comentarios en idioma español considerando lo siguiente:

La importancia de los medios sociales: Las redes sociales se han convertido en un canal de comunicación influyente para compartir información y opiniones. Los movimientos antivacunas han aprovechado estas plataformas para difundir su mensaje y reclutar seguidores. El análisis de medios sociales es esencial para comprender la dinámica de estos movimientos y su impacto en la opinión pública.

Enfoque basado en técnicas de Aprendizaje automático: El aprendizaje automático proporciona herramientas y algoritmos que permiten analizar grandes volúmenes de datos de manera eficiente y extraer patrones y tendencias. Al aplicar técnicas de aprendizaje automático en el análisis de medios sociales, se pueden identificar los principales factores de rechazo de las vacunas, como la propagación de información errónea, la desconfianza en las autoridades sanitarias o los temores relacionados con los efectos secundarios.

Análisis Psicolingüístico: El análisis psicolingüístico se centra en el estudio del lenguaje y su relación con los procesos cognitivos y emocionales. Aplicar técnicas

de análisis psicolingüístico en el estudio de opiniones y comentarios relacionados con los movimientos antivacunas permitirá comprender los aspectos emocionales, las creencias y los sesgos cognitivos presentes en estas discusiones, lo que contribuirá a un mayor entendimiento de los factores que se han ido generando para el rechazo a las vacunas.

Idioma español: Existe una necesidad de investigar y comprender los factores de rechazo de los movimientos antivacunas en el contexto del idioma español, ya que es uno de los idiomas más hablados en el mundo y tiene una gran presencia en las redes sociales. Sin embargo, la mayoría de los estudios sobre movimientos antivacunas se han centrado en el idioma inglés. Éste estudio permitirá obtener información específica y relevante para las comunidades de habla hispana y ayudará a desarrollar estrategias de comunicación y divulgación adaptadas a este público.

Por lo tanto, el desarrollo de un enfoque de análisis de medios sociales basado en técnicas de Aprendizaje automático y análisis Psicolingüístico para determinar los principales factores de rechazo de los movimientos antivacunas en opiniones y comentarios en idioma español es esencial para comprender y abordar eficazmente este problema. El estudio proporcionará información valiosa para la salud pública, las autoridades sanitarias y los profesionales de la comunicación.

1.1.9 Hipótesis

Existe una relación significativa entre las características psicolingüísticas identificadas en los mensajes de los movimientos antivacunas en los medios sociales y los factores de rechazo hacia las vacunas. Al analizar y procesar estos mensajes utilizando técnicas de Procesamiento de Lenguaje Natural y análisis Psicolingüístico, será posible identificar y determinar los principales factores de rechazo que influyen en estos movimientos."

Esta hipótesis sugiere que al analizar los mensajes de los movimientos antivacunas en los medios sociales y aplicar técnicas de Procesamiento de Lenguaje Natural y análisis Psicolingüístico, será posible identificar las características lingüísticas y psicológicas asociadas con el rechazo a las vacunas. Además, sugiere que estas características pueden ser consideradas como factores importantes en el rechazo de los movimientos antivacunas.

1.1.10 Objetivo general

Desarrollar un enfoque de análisis de medios sociales basado en técnicas de Aprendizaje automático y análisis Psicolingüístico para determinar los principales factores de rechazo de los movimientos antivacunas de las opiniones y comentarios en idioma español.

1.1.11 Objetivos específicos

- Realizar un análisis del estado del arte para identificar herramientas de software más utilizadas y las principales técnicas de Procesamiento de Lenguaje Natural y análisis Psicolingüístico.
- Identificar medios sociales tales como como blogs, foros, redes sociales y sitios web de movimientos antivacunas.
- Desarrollar un extractor de información de medios sociales.
- Desarrollar un corpus de opiniones de movimientos antivacunas.
- Realizar la extracción de características psicolingüísticas del conjunto de información obtenida.
- Seleccionar el algoritmo de Aprendizaje automático más apropiado dada la naturaleza de los datos y del problema a resolver.
- Entrenar el algoritmo seleccionado empleando el corpus previamente desarrollado.
- Extraer conocimiento a partir de la explotación de los modelos de predicción resultantes.

1.1.12 Alcances y limitaciones

Alcances:

- **Análisis de medios sociales:** El enfoque propuesto permitirá analizar una gran cantidad de datos provenientes de plataformas de medios sociales como Twitter, Facebook y YouTube lo que brinda una visión amplia de los movimientos antivacunas y sus principales factores de rechazo.
- **Procesamiento de Lenguaje Natural (PLN):** El uso de técnicas de PLN permitirá la extracción y comprensión de características relevantes de los textos en lenguaje natural relacionados con los movimientos antivacunas, lo que facilitará la identificación de los factores de rechazo.
- **Análisis Psicolingüístico:** La aplicación de análisis psicolingüístico, como la evaluación de la polaridad emocional y el tono de los mensajes, proporcionará información adicional sobre los aspectos psicológicos y lingüísticos presentes en los datos de medios sociales relacionados con los movimientos antivacunas.
- **Adaptable a diferentes problemas de clasificación:** El enfoque propuesto podrá adaptarse para abordar otros problemas de clasificación en el ámbito de los medios sociales y el lenguaje natural, lo que amplía su aplicabilidad más allá de los movimientos antivacunas.

Limitaciones:

- **Disponibilidad y calidad de los datos:** La efectividad del enfoque dependerá de la disponibilidad y calidad de los datos recopilados de los medios sociales. Si los datos son limitados o sesgados, podría afectar la precisión de los resultados obtenidos.
- **Variabilidad y evolución de los movimientos antivacunas:** Los movimientos antivacunas son dinámicos y evolucionan con el tiempo. El enfoque propuesto podría no capturar todas las variaciones y cambios en los factores de rechazo a medida que evolucionan los movimientos.
- **Idioma y contexto:** El enfoque se centrará en el análisis de medios sociales en el idioma español y en el contexto específico de los movimientos antivacunas. Esto limita su aplicabilidad a otros idiomas y contextos culturales.
- **Generalización de los resultados:** Los resultados obtenidos podrían ser específicos de los movimientos antivacunas estudiados y no necesariamente se pueden generalizar a otros grupos o contextos de rechazo a las vacunas.

CAPÍTULO II ESTADO DEL ARTE

2.1.1 Trabajos relacionados

Existen varias investigaciones que se han realizado con base en las plataformas de las redes sociales donde se libra una batalla entre aquellos que sustentan bajo términos científicos sobre los beneficios de la aplicación de una vacuna y también lo que se opina y argumenta de forma contraria o negativa.

Un grupo de científicos en Japón (Shi et al., 2022) realizaron una investigación de las características psicolingüísticas de los antivacunas en la red social Twitter. En este estudio recopilaron Tweets relacionados con COVID-19 de febrero de 2020 a junio de 2021 con el objetivo de realizar un análisis de la postura de vacunación en la población de este medio en el cual se consideraron los rasgos lingüísticos y las características propias de la red social.

Se utilizó API Twitter para recopilar tweets, respuestas y retweets (RT) relacionados con COVID-19. La recolección de datos fue enfocada en tweets en idioma inglés. Se utilizó el algoritmo de Louvain para la detección de comunidades que están a favor y en contra de las vacunas. Se tomo de apoyo el software Ghepi para revisar la estructura de los clústeres resultantes. La red de retuits fue cuantificada de acuerdo a su densidad, coeficiente de agrupación y la distancia media. Adicional fue utilizada la herramienta estandarizada en psicología social llamada LIWC para medir los sentimientos y los estilos de pensamiento en un tuit. Se utilizó el diccionario de fundamentos morales (MDF) de la versión japonesa para medir cinco fundamentos morales como características psicolingüísticas adicionales para los grupos pro y antivacunas en Twitter.

Los resultados obtenidos demostraron que al realizar una comparación de los que están a favor de las vacunas con aquellos que están en contra, estos últimos tienden a tener más emociones negativas, pensamiento narrativo, y tendencias inmorales.

Jae-Geum Shim y varios colaboradores (Shim et al., 2021) efectuaron un estudio sobre la respuesta pública a las vacunas contra COVID-19 en las redes sociales. La información recopilada fueron tweets del 23 de febrero al 22 de marzo de 2021. Se llevo a cabo un estudio de este tema para interpretar y hacer una clasificación de los tuits. Fue utilizado el modelo Latent Dirichlet Allocation (LDA) el cual es utilizado en el procesamiento de lenguaje natural para buscar textos similares. Adicional se llevó a cabo un análisis de sentimientos para poder examinar las emociones del público usando el modelo. El resultado de este análisis de sentimientos arrojó una proporción similar en opiniones positivas y negativas antes del inicio de las vacunas. Sin embargo, aumentaron las opiniones negativas después del número de casos confirmados de COVID-19. Se puede observar que la anticipación, la decepción y el miedo es lo que se refleja en los tuits.

Nihal Durmaz y Engin Hengirmen (Durmaz & Hengirmen, 2022) Crearon un estudio sobre la red social Twitter en la cual se identificaron interacciones comunitarias que estuvieran relacionadas con las vacunas y se identificaron influencias del antes y después de la pandemia de COVID-19. Las muestras de recopilación de datos utilizadas fueron categorizadas en dos grupos (antes y después de la pandemia) en períodos del 1 de junio de 2019 al 10 de marzo de 2020 y del 11 de marzo de 2020 al 1 de enero de 2022. Estos conjuntos de datos se recopilaron con Twint, un código abierto de Python que permitió recopilar datos sin restricciones en Twitter.

Fue creada una red social a partir de tuits, primero se realizó una depuración de tweets irrelevantes. Se creó código Python para obtener una lista de todos los hashtags utilizados en los tuits. Se compilaron un total de 8393 hashtags para el período anterior y 29956 hashtags para el período posterior. Después de filtrar los tweets relevantes se movieron a un archivo CSV, que Gephi puede leer como una red social. Fue usado SNA (análisis de redes sociales) para entender la sensibilidad de la vacuna en la población. Este estudio investigó el discurso actual sobre vacunas en Twitter, examinó las opiniones de la sociedad sobre las vacunas,

e identificó a los actores clave. Los resultados arrojados en la red pre-COVID-19, las figuras de los medios y los autores que tenían puntos de vista antivacunas fueron los usuarios más influyentes. En la red post-COVID-19, el ministro de salud turco, fue la figura más influyente. Se observó que la red de vacunas estaba creciendo rápidamente después de COVID-19, y las opiniones de médicos sobre las vacunas obligatorias recibieron una gran reacción. Este estudio comprobó la efectividad y utilidad que se puede obtener de las redes sociales para comprender la opinión sobre la salud pública y la vacunación en Turquía y el intento de reducir la infodemia antes de administrar la vacuna y apoyar a la Salud Pública para actuar de forma proactiva ante el conocimiento obtenido.

En España fueron recopilados tres estudios ligados a la investigación de los movimientos a favor y en contra de las vacunas. El primero de ellos (Prieto Santamaría et al., 2022) se enfocó en las redes sociales, y muy particular en Twitter; para mostrar y analizar el impacto y las tendencias de los mensajes publicados en español que hicieron referencia a influenza, sarampión y MMR del 1 de enero de 2015 al 31 de diciembre de 2018. La metodología que fue utilizada consistió en realizar una selección y extracción de la muestra a estudiar. Para la extracción de los tuits se manejaron herramientas disponibles en Twitter (utilizaron la API oficial - Interfaz de programación de aplicaciones). Fueron identificados los sentimientos expresados en los mensajes utilizando herramientas comerciales y también fueron desarrollados nuevos mecanismos basados en aprendizaje automático (ML) y un análisis final descriptivo de los resultados obtenidos. Como resultado el análisis arroja que los tweets relacionados con la influenza, sarampión y la MMR han demostrado que existe un dilema heredado de la población cuando se trata de las vacunas y la aplicación de las mismas.

El segundo estudio (Conlledo et al., 2020) elaboró un artículo sobre los argumentos en pro de las vacunas y los antivacunas que fueron compartidos en blogs, Facebook y YouTube mediante la recopilación de publicaciones realizadas desde el año 2015 a 2022. En este análisis se consideró la fuente origen de la

información, el contenido y el impacto creado. En la muestra de estudio se consideraron 48 videos de YouTube, 207 entradas de 5 blogs distintos con tendencias dispares y 7 grupos de carácter público de Facebook. Los resultados obtenidos mostraron que la información a favor y en contra no proviene de organismos oficiales, el contenido que fue compartido en ciertos temas estaba ligado a hechos puntuales que les sucedieron de forma particular, normalmente el contenido carecía de carácter científico y las interacciones entre usuarios variaba de acuerdo si la información transmitida era por YouTube, Facebook o un blog.

El último estudio (Larrondo-Ureta et al., 2021) consistió en medir y caracterizar la información sobre las vacunas contra el Covid-19 en la red social Twitter. Esto fue considerado en un contexto de infodemia del cual parten todo tipo de contenido de información y desinformación además de la falsedad sobre la vacuna ante la crisis sanitaria desatada por la pandemia de Covid-19. Fueron analizados 62.045 tuits y 258.843 retuits de todos aquellos que estuvieron a favor y en contra de la vacuna entre diciembre de 2020 y febrero de 2021. Se puede decir que cuando se inició la campaña de vacunación fue la punta de lanza que desató que los discursos provacuna iniciaran con una gran mayoría ante los que estaban en contra. Pero los grupos antivacunas se destacaron por ser clústeres fuertemente cohesionados y un gran nivel de actividad, pero con un impacto menor en viralizar contenidos. Se concluye que el apoyo de los discursos en contra de la vacuna lo hacen en base a medios alternativos o contenidos que se comparten en las redes sociales, por lo tanto, se confirma que la información con contenido de calidad es una base contundente contra la desinformación.

Krzysztof , Agnieszka, Karolina y Łukasz BalwickiE (Klimiuk et al., 2021) Realizaron una exploración y estudio el cual tuvo como objetivo: "utilizar las redes sociales para identificar y analizar los argumentos de todos aquellos que niegan las vacunas". Fueron extraídos de Facebook todos los comentarios de oposición a la vacuna de COVID-19 entre el 05 de enero al 31 de Julio de 2019. El método

utilizado fue la recopilación de publicaciones y comentarios de la página de Facebook más popular que aboga por la negativa de las vacunas obligatorias en Polonia. Los datos se ensamblaron en tres hojas de cálculo separadas (una para cada mes). Los comentarios fueron clasificados con base en Kata, quien dividió el contenido antivacunación en categorías. Todos los comentarios recogidos en el estudio fueron analizados con la herramienta de software Sentimenti para el análisis de sentimientos de texto. Se crearon estadísticas descriptivas para analizar los argumentos en las publicaciones y comentarios. El resultado obtenido fue que muchos de los comentarios contenían teorías conspirativas, en otros, prevalecía la desinformación y premisas poco fiables, algunos fueron enfocados a la seguridad y efectividad de las vacunas, otros fueron clasificadas en cuestión de moralidad, religión y creencias. También hubo una categoría de comentarios de burla y ofensivos. En conclusión, el análisis de sentimientos arrojó que comentarios sin argumentos y todos aquellos que contenían sugerencias sobre medicina alternativa o desinformación fueron de alguna forma más positivos y con menos enojo que los comentarios hechos a otras categorías.

En el año 2015 (California Disneylandia) después de un rebrote de sarampión Xiaoyi Yuan y Andrew T. Crooks (Yuan & Crooks, 2018) realizan una investigación en el que se trabajan los patrones de comunicación de los usuarios que están a favor de la vacunación (provacunas) y aquellos que están en contra (antivacunas) y el papel de los bots en Twitter en base al análisis de una red de Re tweets relacionados con la vacuna MMR (vacuna triple vírica -- sarampión, paperas y rubéola). La población de estudio fue recopilada del 1 de febrero al 9 de marzo de 2015. Esta muestra fue considerada en este lapso de tiempo ya que el primer caso identificado fue el 5 de enero de 2015 y fueron notificados casos hasta finales de 2015. Fue elegido este periodo de análisis porque el evento provocó un gran volumen de discusión sobre la vacuna MMR en línea. En esta investigación se utiliza una combinación de análisis de sentimientos con aprendizaje automático y la detección de comunidades en las redes sociales para encontrar los patrones de

usuario que estuvieron a favor y en contra de la vacuna. Esta investigación descubre que las opiniones de los usuarios antivacunas están muy segregadas de los usuarios con opiniones a favor de la vacunación y que las comunidades antivacunas residen en su propia comunidad estructural cerrada, lo cual significa que retuitear se da con mucha más frecuencia dentro de sus propios grupos de opinión que entre grupos diversos.

Augusto Rodríguez Orejuela junto con 2 colaboradores (Rodríguez Orejuela et al., 2021) Realizan una investigación la cual consistió en analizar los sentimientos subyacentes sobre la vacuna contra COVID-19 en la red social Twitter. La cantidad de comentarios recopilados de Twitter desde el 12 de mayo hasta el 30 de septiembre del 2020 fue de 38.034. El procedimiento utilizado fue recopilar las publicaciones mediante minería de datos y aplicación de técnicas Machine Learning, análisis de sentimientos y análisis de redes para poder identificar los sentimientos expresados por los usuarios a favor y en contra de la campaña de vacunación contra de COVID-19. El resultado obtenido fue que los sentimientos ante la vacunación en su gran mayoría fueron negativos. Las emociones identificadas como más recurrentes son el miedo y la ira. Sin embargo, esto no quiere decir que el volumen de personas era mayormente de las que rechazaban la vacunas como medida de protección, pero sí se puede decir que el producto resultante del escenario desesperanzador de la pandemia incentivó la incertidumbre y el miedo, dejando con poco porcentaje al optimismo.

2.1.2 Análisis comparativo de los trabajos relacionados.

Tabla 2.1 Análisis comparativo de trabajos relacionados.
Fuente: Propia.

Título	Autores	Fecha Publicación	¿Qué investigaron?	Muestra	Metodología	Resultados
Psycho-linguistic differences among competing vaccination communities on social media https://doi.org/10.48550/arxiv.2111.05237	Jialiang Shi, Piyush Ghasiya, and Kazutoshi Sasahara	18/03/2022	Se analizaron las características psicolingüísticas de los antivacunas en la red social Twitter.	Tweets relacionados con COVID-19 de febrero de 2020 a junio de 2021	** Recopilación de datos. *** Caracterización de grupos anti y pro vacunas. ** Medidas de red. ** Análisis psicolingüístico con LIWC. ** Análisis de moralidad con MFD.	Los antivacunas tienden a tener más emociones negativas en comparación con los que están a favor de las mismas.
Text Mining Approaches to Analyze Public Sentiment Changes Regarding COVID-19 Vaccines on Social Media in Korea https://pubmed.ncbi.nlm.nih.gov/34207016/ DOI: 10.3390/ijerph18126549	Jae-Geum Shim, Kyoung-Ho Ryu, Sung Hyun Lee, Eun-Ah Cho, Yoon Ju Lee and Jin Hee Ahn	18/06/2021	Se investigó la respuesta pública coreana a las vacunas COVID-19 en las redes sociales	Recopilación de tweets relacionados con las vacunas COVID-19 del 23 de febrero de 2021 al 22 de marzo de 2021.	** Fuente de datos y preprocesamiento. ** Corpus de tweets utilizando un modelo de tema probabilístico. ** Modelado de temas y LDA (Latent Dirichlet Allocation). ** Análisis de sentimientos ** Análisis estadístico.	Se considera que la anticipación, la decepción y el miedo con respecto a las vacunas se reflejan en los tuits.
The dramatic increase in anti-vaccine discourses during the COVID-19 pandemic: a social network analysis of Twitter https://pubmed.ncbi.nlm.nih.gov/35113767/ DOI: 10.1080/21645515.2021.2025008	Nihal Durmaz a and Engin Hengirmen	03/02/2022	El objetivo de este estudio fue identificar interacciones comunitarias relacionadas con las vacunas e identificar influencias clave/personas influyentes antes y después de la pandemia usando datos de la red social de Twitter.	Fueron considerados dos conjuntos de datos * 1 de junio de 2019 al 11 de marzo de 2020. * 11 de marzo de 2020 al 1 de enero de 2021.	** Consideraciones éticas. ** Conjunto de datos. ** Preparación de datos. ** Se usa SNA para entender la red de vacunación. ** Construcción de redes y visualización.	Este estudio muestra la efectividad y utilidad de los datos de redes sociales para comprender la opinión pública sobre la salud y la vacunación en Turquía.
Influenza and Measles-MMR: two case study of the trend and impact of vaccine-related Twitter posts in Spanish during 2015-2018 https://pubmed.ncbi.nlm.nih.gov/33662222/ DOI: 10.1080/21645515.2021.1877997	Lucia Prieto Santamaria, Juan Manuel Tuñas, Diego Fernández Peces-Barba, Almudena Jaramillo, Manuel Cotarelo, Cecilia Martínez	04/03/2021	El objetivo del trabajo es mostrar y analizar el impacto y las tendencias de los mensajes publicados en Twitter en español referidos a influenza, sarampión y MMR.	Fueron considerados los mensajes relacionados con las vacunas publicados en español durante 2015-2018.	** Selección de muestras. ** Extracción de muestras. ** Análisis de sentimientos. ** Glosario de términos.	Los resultados mostraron que, contrario a la opinión popular, la mayoría de los mensajes publicados no fueron negativos.
Análisis del contenido publicado en YouTube, Facebook e internet sobre vacunas y antivacunas https://doi.org/10.35669/rccs.2020.10(1).67-90	Paloma Piqueiras Conlledo, Giorgio De Marchis, Victoria Cuesta Diaz	06/08/2020	El objetivo general de esta investigación consiste en analizar el tipo de información relacionada con la vacunación que se distribuye en blogs, YouTube, y grupos de Facebook.	Fueron considerados para su estudio 48 videos de YouTube, 207 post de 5 blogs distintos y de 7 grupos de Facebook. Los datos fueron recopilados entre el 23 de diciembre de 2019 y el 23 de enero de 2020.	Desarrollada en dos pasos: a) Se estipuló un procedimiento exhaustivo de registro de los datos. b) Se definieron las especificidades de cada campo de búsqueda. ** Procedimiento de registro de los datos. ** Explicación de los campos registrados en cada canal. ** Muestra.	Se encuentra que: a) anti vacuna o pro vacuna no suele provenir de organismo oficiales. b) el contenido gira en torno a temas recurrentes. c) el contenido carece de carácter científico y; d) las interacciones varían mucho dependiendo de si la información es transmitida por YouTube, un blog, o Facebook.
Desinformación, vacunas y Covid-19. Análisis de la infodemia y la conversación digital en Twitter DOI: https://doi.org/10.4185/RLCS-2021-1504	Ainara Larrondo-Ureta, Simón-Peña Fernández, Jordi Morales-I-Gras	15/04/2021	Este artículo tuvo como finalidad medir y caracterizar la conversación sobre las vacunas contra el Covid-19 en la red social Twitter.	Fueron filtrados y considerados 62.045 tuits y 258.843 retuits de partidarios y detractores de la vacuna entre diciembre de 2020 y febrero de 2021	** Descarga datos de conversación ** Algoritmo de modularidad para identificar clústeres. ** Partición de red	Los grupos antivacunas se caracterizan por ser clústeres fuertemente cohesionados, con un apreciable nivel de actividad, pero con menor capacidad para viralizar contenidos.
Vaccine misinformation on social media – topicbased content and sentiment analysis of Polish vaccine-deniers' comments on Facebook DOI: 10.1080/21645515.2020.1850072	Krzysztof Klimiuk, Agnieszka Czaska, Karolina Biernacka & Lukasz Balwicki	30/01/2021	Este estudio tuvo como objetivo utilizar las redes sociales para identificar y analizar los argumentos de los que niegan las vacunas.	Se consideraron todos los comentarios publicados en la página de Facebook (de uno de los principales opositores a la vacunación de Polonia) entre el 05/01/2019 y el 31/07/2019.	** Recopilación de datos. ** Análisis de Sentimientos. ** Análisis Estadístico. Todo los datos fueron recopilados y analizados cuantitativamente en cuanto a su contenido.	Los autores de comentarios se caracterizaron por una falta de confianza en los logros científicos de la medicina. Estos hallazgos deben abordarse adecuadamente en las campañas de vacunación.
Examining Online Vaccination Discussion and Communities in Twitter https://doi.org/10.1145/3217804.3217912	Xiaoyi Yuan, Andrew T. Crooks	18/07/2018	El propósito radicó en investigar los patrones comunicativos de los usuarios antivacunas y provacunas en Twitter mediante el estudio de la red de retuits.	de 660 892 tuits relacionados con la vacuna MMR publicados por 269 623 usuarios después del brote de sarampión de 2015 en California Disneyland.	** Recolección de datos ** Preproceso ** Extracción de características ** Entrenamiento de múltiples clasificadores con los datos etiquetados. ** Construcción red de retuits. ** Detección de comunidad.	El documento concluye que los usuarios de Twitter antivacunas son comunidades agrupadas y cerradas, y esto dificulta que las organizaciones de salud penetren y contrarresten la información obstinada.
Sentimientos hacia la vacunación contra la covid-19: panorama colombiano en Twitter DOI:10.5294/pacla.2022.25.1.4	Augusto Rodríguez-Orejuela, Claudia Lorena Montes-Mora, Carlos Fernando Osorio-Andrade	17/12/2021	Esta investigación consistió en analizar los sentimientos subyacentes sobre la vacuna contra COVID-19 en la red social Twitter.	La cantidad de comentarios recopilados de Twitter abarcan desde el 12 de mayo hasta el 30 de septiembre del 2020.	** Análisis de sentimientos en redes sociales *** Selección de palabras clave *** Extracción de datos *** Procesamiento del texto *** Aprendizaje de máquina *** Evaluación de algoritmos *** Pesado de características *** Grafos de redes sociales	Los resultados sugieren que, en su mayoría, los sentimientos hacia las vacunas son negativos. El miedo y la ira, respectivamente, son las emociones más recurrentes en las opiniones de Twitter.

2.1.3 Propuesta de solución

4.1.3.1 Análisis de las Tecnologías de Información

Atendiendo el propósito de investigación y estudio de este proyecto, y considerando las bondades de la minería de datos que permite comprender y filtrar grandes cantidades de datos; se consideró realizar un estudio previo de las herramientas tecnológicas (con código abierto) que pueden ser utilizadas para el desarrollo de este trabajo. La **Tabla 2.2** muestra las opciones tecnológicas que pueden ser ocupadas en esta investigación.

Tabla 2.2 Alternativas tecnológicas de solución Fuente: Propia

Tecnología	Componentes	
Python Permite la limpieza y organización de datos hasta la aplicación de algoritmos de aprendizaje automático	Visualizar Datos	Matplotlib, Seaborn Bokeh
	Analizar datos	NumPy, SciPy, Pandas, Numba
	Aprendizaje Automatico	scikit-learn
	Aprendizaje	TensorFlow, Keras
	IA	SHAP
	Procesamiento de Lenguaje	NLTK: Natural Language Toolkit gensim, spaCy
R Permite la programación para análisis estadístico y puede trabajar con grandes volúmenes de datos	Manipulación de Datos	dplyr, data.table, lubridate, jsonlite
	Despligie de Gráficos	ggplot2, Corrplot, lattice.
	Aprendizaje Automatico	mlr, XGBoost, caret,gbm,Prophet, randomforest
	Investigación Reproducible	Slidify

4.1.3.2 Tecnologías Seleccionadas

Python

Es un lenguaje de programación de alto nivel, interpretado y generalmente considerado como un lenguaje de propósito general. Es ampliamente utilizado en la minería de datos debido a sus grandes bondades y ventajas que ofrece en esta tarea. Algunas de las características más relevantes de esta tecnología son:

- **Facilidad de uso:** Es conocido por su sintaxis sencilla y legible, lo que lo convierte en un lenguaje fácil de aprender y entender. Esto facilita la implementación de algoritmos y el desarrollo de soluciones en minería de datos.
- **Amplias bibliotecas y herramientas:** Cuenta con una gran cantidad de bibliotecas y herramientas diseñadas específicamente para el análisis de datos y la minería de datos. Algunas de las bibliotecas más populares incluyen NumPy, pandas, scikit-learn, TensorFlow y PyTorch. Estas bibliotecas proporcionan una amplia gama de funciones y algoritmos para tareas de minería de datos, como clasificación, clustering, regresión y procesamiento de lenguaje natural.
- **Comunidad activa:** Tiene una comunidad de desarrolladores muy activa y comprometida, lo que significa que hay una gran cantidad de recursos disponibles, como documentación, tutoriales y ejemplos de código. Además, existe un ecosistema robusto de paquetes y módulos desarrollados por la comunidad que facilitan el trabajo en minería de datos.
- **Integración con otras tecnologías:** Se integra fácilmente con otras tecnologías y herramientas utilizadas en la minería de datos.
- **Escalabilidad:** Se distingue por su escalabilidad, lo que significa que puede manejar grandes volúmenes de datos de manera eficiente. Además, es compatible con el procesamiento en paralelo y la computación distribuida, lo que permite realizar análisis de datos a gran escala.

LIWC

Es una herramienta ampliamente utilizada en el análisis de texto y la investigación en áreas como la psicología, la lingüística, la comunicación y la minería de datos. Aquí hay algunas razones por las que se utiliza LIWC en este campo:

- **Análisis lingüístico:** Permite realizar un análisis lingüístico detallado del texto. Utiliza un diccionario que asigna palabras a diferentes categorías lingüísticas, como emociones, cogniciones, categorías gramaticales y estilos de procesamiento

cognitivo. Esto proporciona información valiosa sobre el contenido emocional, cognitivo y lingüístico del texto.

- **Automatización y eficiencia:** Automatiza el análisis lingüístico al asignar automáticamente las palabras del texto a las categorías correspondientes. Esto ahorra tiempo y esfuerzo en comparación con el análisis manual de grandes volúmenes de texto. Es capaz de procesar rápidamente grandes cantidades de texto, lo que lo convierte en una herramienta eficiente en la minería de datos.
- **Extracción de características:** Proporciona un conjunto de características lingüísticas cuantitativas que se pueden utilizar como variables para realizar análisis estadísticos y modelos predictivos. Estas características pueden incluir recuentos de palabras en diferentes categorías lingüísticas, proporciones de palabras emocionales y cognitivas, y medidas de estilo de procesamiento cognitivo.
- **Validación y comparabilidad:** Ha sido ampliamente utilizado y validado en numerosos estudios de investigación en diferentes campos, lo que lo hace confiable y permite comparaciones entre diferentes estudios. Su uso generalizado en la comunidad científica proporciona una base sólida para la interpretación y comparación de resultados.
- **Aplicaciones en diferentes dominios:** Se ha utilizado en una amplia gama de dominios de investigación, incluyendo psicología, ciencias sociales, salud, análisis de sentimientos, minería de opiniones y más. Su flexibilidad y adaptabilidad lo convierten en una herramienta versátil para el análisis de texto en diferentes contextos.

WEKA

Es una suite de software en la minería de datos y aprendizaje automático con flexibilidad de código abierto. Proporciona un conjunto completo de herramientas para realizar tareas como clasificación, regresión, agrupación, asociación, selección de atributos y visualización de datos. Algunas de sus grandes ventajas de uso son las siguientes:

- **Amplia gama de algoritmos:** Ofrece una amplia variedad de algoritmos de aprendizaje automático y técnicas de minería de datos. Esto incluye algoritmos para clasificación, regresión, agrupación, asociación, extracción de reglas y más. La disponibilidad de una amplia gama de algoritmos permite probar y comparar diferentes técnicas para encontrar la más adecuada para un problema específico.
- **Interfaz gráfica intuitiva:** Proporciona una interfaz gráfica de usuario fácil de usar, lo que hace que sea accesible tanto para principiantes como para usuarios experimentados. La interfaz gráfica permite realizar tareas comunes de minería de

datos, como cargar datos, preprocesarlos, aplicar algoritmos y visualizar resultados, de manera visual e interactiva.

- **Flexibilidad y extensibilidad:** Es altamente flexible y extensible. Permite implementar y probar sus propios algoritmos y técnicas personalizadas utilizando Java. Esto brinda la capacidad de adaptar Weka a sus necesidades específicas y explorar nuevas ideas en minería de datos.
- **Evaluación y validación de modelos:** Proporciona herramientas y métodos para evaluar y validar los modelos de aprendizaje automático generados. Permite realizar pruebas cruzadas, división de datos de entrenamiento y prueba, y realizar diversas métricas de evaluación, como precisión, sensibilidad, especificidad, entre otras. Esto facilita la evaluación y comparación de diferentes modelos y técnicas.
- **Documentación y comunidad activa:** Cuenta con una extensa documentación en línea, que incluye tutoriales, ejemplos de uso y una guía del usuario. Además, tiene una comunidad activa de usuarios y desarrolladores que brindan soporte, comparten conocimientos y contribuyen con nuevas funcionalidades y mejoras.

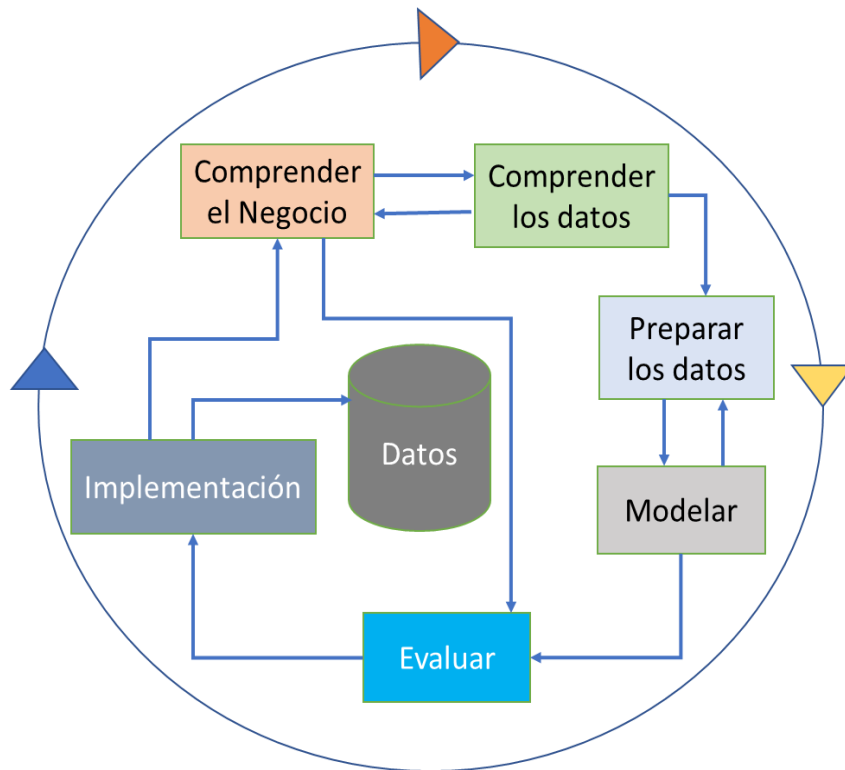
CAPÍTULO III METODOLOGÍA Y DESARROLLO

3.1.1 Metodología de investigación/desarrollo

De acuerdo al desarrollo de investigación planteado en este trabajo de tesis se establece utilizar el modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), técnica probada para orientar aquellos trabajos en minería de datos.

- Como metodología, proporciona explicaciones sobre las etapas típicas de un proyecto, las actividades requeridas en cada etapa y una comprensión de las conexiones entre dichas actividades.
- Como modelo de proceso, CRISP-DM presenta una visión general del ciclo de vida de la minería de datos (IBM, 2021) .

Al ser tan flexible permite que este proyecto pueda adaptarse a la estrategia planteada, ya que se requiere proporcione una estructura sólida para abordar los desafíos de la minería de datos aquí presentados y garantizar que las tareas realizadas puedan ser llevadas a cabo de manera efectiva y eficiente. La **Figura 3.1** representa las etapas principales y el flujo de proceso de la metodología CRISP-DM



*Figura 3.1 Representación de Metodología CRISP-DM
Fuente: Propia.*

Siguiendo cada uno de los pasos establecidos a continuación se detallan las actividades realizadas en cada uno de ellos hasta conseguir la obtención de resultados.

3.1.2 Comprensión del negocio

Como ya ha sido mencionado anteriormente el objetivo de esta tesis es proporcionar una comprensión profunda de los factores que impulsan el rechazo a las vacunas, identificando los patrones lingüísticos, las emociones y los argumentos asociados con estos movimientos a través del análisis de medios sociales. Los resultados obtenidos pueden ayudar a informar estrategias de comunicación, educación y divulgación que aborden las preocupaciones y mitos sobre las vacunas, promoviendo así una mayor aceptación y confianza en la inmunización.

Por consiguiente, los puntos que serán abordados en este trabajo son:

- Evaluar la polaridad de las palabras utilizadas en los textos para revelar si están expresando sentimientos positivos o negativos hacia las vacunas. Palabras con connotaciones negativas o críticas que puedan indicar un rechazo hacia las vacunas.
- Identificar palabras emocionales utilizadas en los textos, como términos relacionados con el miedo, la desconfianza o la indignación, que puedan revelar las emociones asociadas con el rechazo de las vacunas.
- Analizar el estilo de argumentación utilizado en los textos que puedan brindar información sobre los razonamientos y fundamentos detrás del rechazo de las vacunas. Por ejemplo, se pueden identificar argumentos basados en teorías de conspiración, desinformación o temores sobre la seguridad de las vacunas.
- Observar el uso de términos y frases específicas relacionadas con los movimientos antivacunas que puedan ayudar a identificar los textos que promueven o apoyan el rechazo de las vacunas.
- Analizar la estructura discursiva de los textos, como la presencia de afirmaciones categóricas, negaciones o contradicciones que pueda proporcionar conocimientos significativos sobre la argumentación y la postura de rechazo hacia las vacunas.

3.1.3 Comprensión de los datos

En esta etapa, se han considerado los medios sociales Twitter, Facebook y Youtube para recopilar y explorar los datos relacionados con los movimientos antivacunas que puedan estar disponibles y ser de ayuda para el trabajo de esta investigación (Siempre respetando las políticas y regulaciones establecidas).

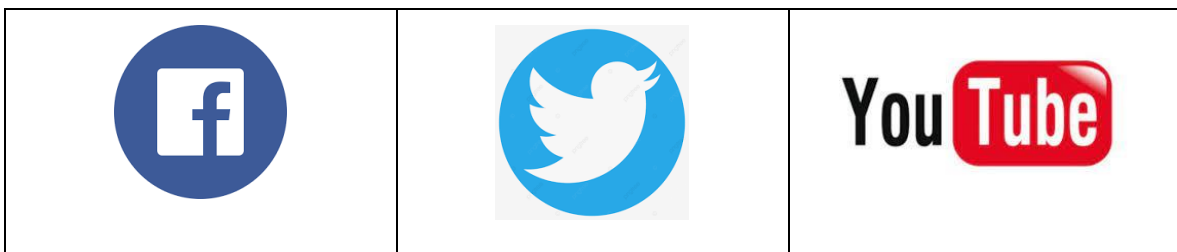


Figura ## Logotipos oficiales tomados de internet

Algunas de las tareas realizadas en esta fase incluyen:

4.1.3.1 Exploración y análisis preliminar de los datos

Aquí se realizó una exploración inicial de los datos que pueden ser recopilados para comprender su estructura y características. Esto con el fin de poder identificar variables relevantes, la visualización de patrones o tendencias iniciales, y la comprensión de la distribución de los datos. De la **Figura 3.3 a la Figura 3.8** se presentan una extracción de los comentarios obtenidos de los medios sociales y su tendencia o clasificación respecto a una posición de opinión.

```
{ ID:1622904576104972294, text: @LauLaLokita Mi señora madre es anti vacunas,  
fan de las teorías conspirativas, dice que el covid no existe y que todo es un invento  
del gobierno que nos quiere matar y mi papá un victimista con problemas de ira infiel,  
golpeador y anti vacunas tambien 🙄🙄🙄
```

Figura 3.3 TWITT Negativo sobre vacuna Fuente: Twitter.

```
{ ID:1623516821046325248, text: RT @Mippcivzla: La vacuna anti-SARS-CoV-2 permitiría generar inmunidad comunitaria.  
Haciendo que la población disminuya probabilidades de contagio 🙄
```

Figura 3.4 TWITT Positivo sobre vacuna Fuente: Twitter.

Confirman caso de sarampión en Monterrey
Lamentable este hecho, 🙄🙄
No se olviden de vacunar, todos estamos en riesgo!!
#provacunas #cartilladevacunacionaldia

Figura 3.5 Comentario Negativo FB sobre vacuna Fuente: Facebook

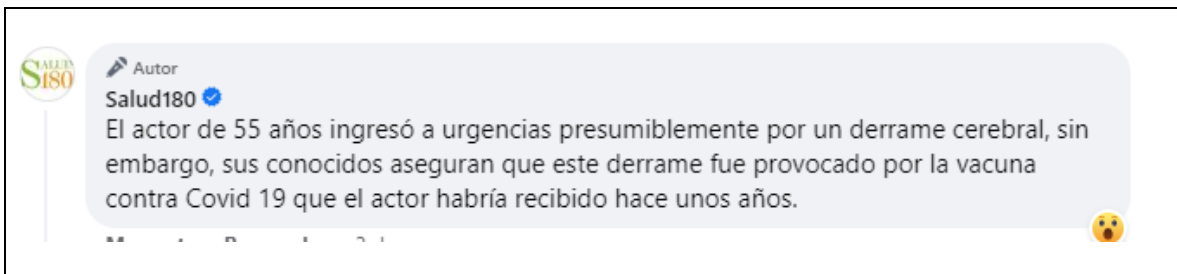


Figura 3.6 Comentario Positivo FB sobre vacuna Fuente: Facebook.

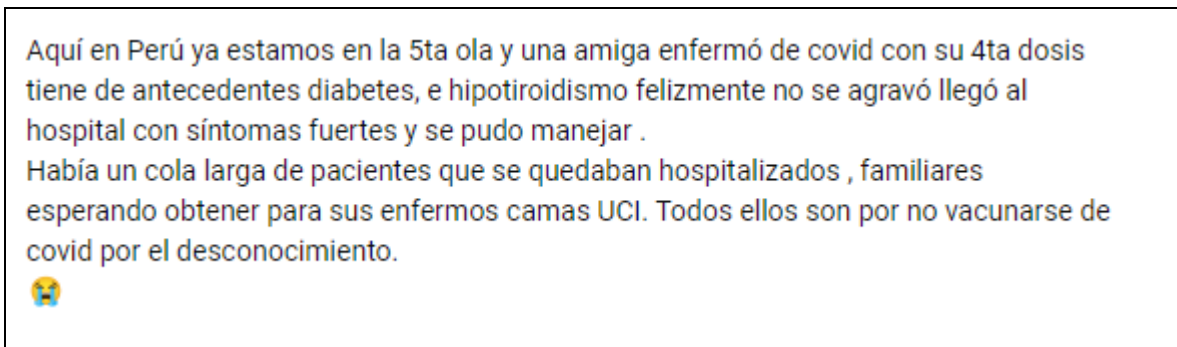


Figura 3.7 Comentario Negativo YouTube referente a video de Vacunas Fuente: YouTube.

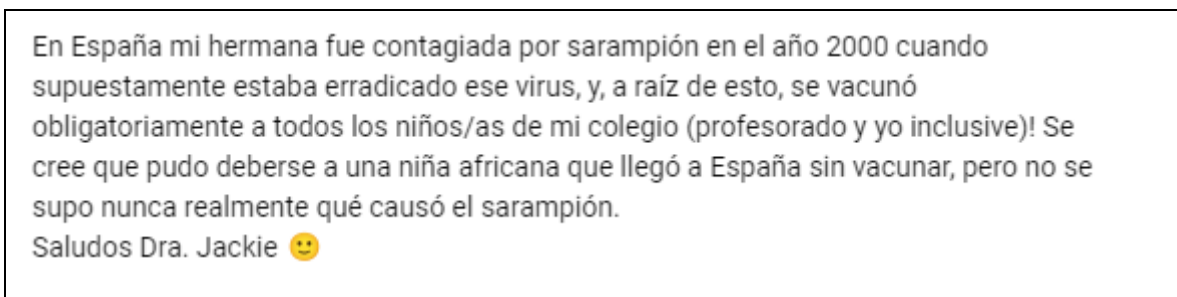


Figura 3.8 Comentario Positivo YouTube referente a video de Vacunas Fuente: YouTube

4.1.3.2 Análisis de contenido

Se llevó a cabo un análisis de contenido para comprender el significado y la temática de los mensajes y publicaciones relacionados con los movimientos antivacunas. Esto podría implicar identificar palabras clave, conceptos recurrentes, temas principales

y subtemas dentro de los datos. Además, se podrían explorar las opiniones, actitudes y argumentos expresados en los textos relacionados con la vacunación, vacuna, movimientos antivacunas, provacunas y todo lo relacionado en favor o en contra de la aplicación de las vacunas.

Ya que el objetivo principal es proporcionar una comprensión profunda de los factores que impulsan el rechazo a las vacunas, identificando los patrones lingüísticos, las emociones y los argumentos asociados con estos movimientos a través del análisis de los medios sociales se identifica un Corpus de datos de los diferentes medios de (archivos de texto con contenido de tweets y/o comentarios). En la **Tabla 3.1** se pueden observar el número de Total de datos obtenidos para el análisis de los datos.

Tabla 3.1 Cometarios por cada medio social de estudio Fuente: Propia.

	Total de Datos
Twitter	3000
FB	2000
YouTube	1500

3.1.4 Preparación de los datos

En esta fase, se llevó a cabo el procesamiento y la transformación de los datos para que sean adecuados para su análisis posterior. Esto incluye la limpieza de datos, la selección de características relevantes, la codificación de variables categóricas, la normalización de datos y la creación de nuevas variables derivadas.

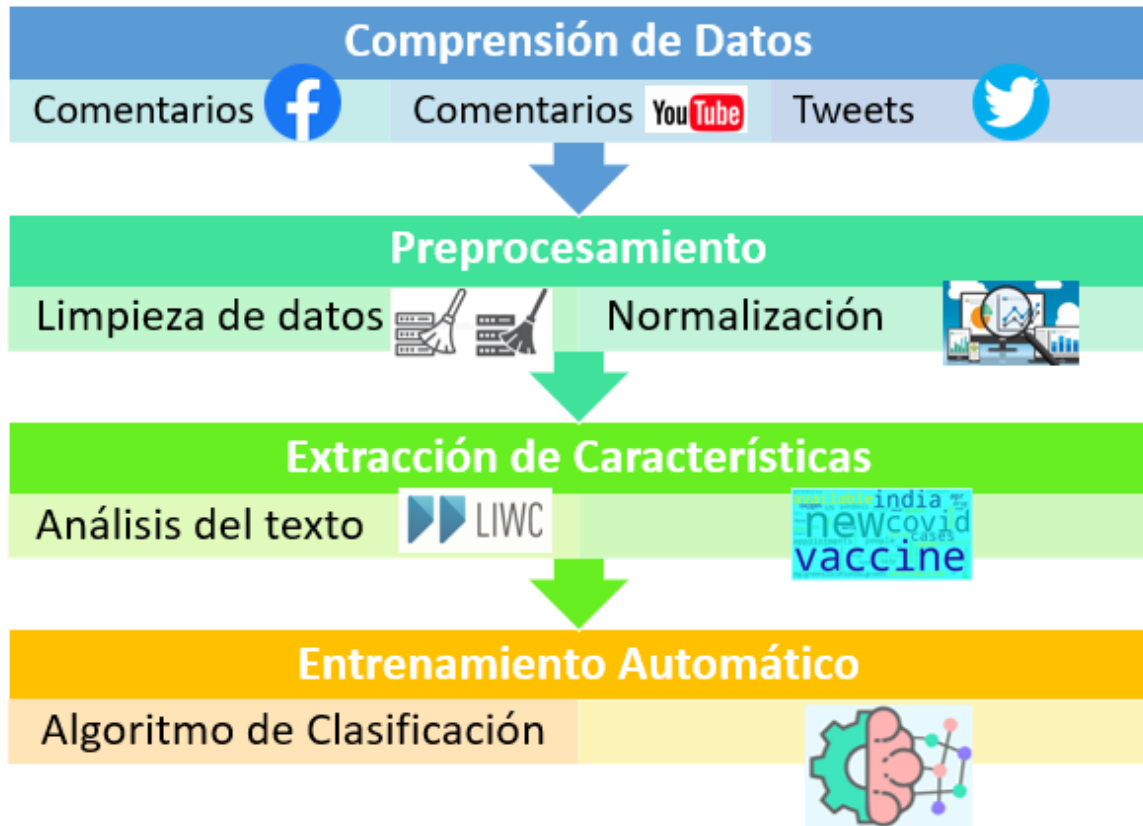


Figura 3.9 Propuesta de Clasificación paso a paso Fuente: YouTube

Algunas de las actividades que se llevarían a cabo en esta fase incluyen:

4.1.4.1 Recopilación de datos

En esta fase se realizó la tarea de recopilar los datos relevantes de los medios sociales Twitter, Facebook y YouTube que hacen referencia a los movimientos antivacunas. Esto implicó la extracción de mensajes, publicaciones, comentarios u otros tipos de contenido relacionado con el tema.

- **Cuenta desarrollador Twitter:** Para acceder a la API de Twitter fue requerida una cuenta de desarrollador. Se realizó registro en el sitio web de Twitter Developer y se llevó a cabo el proceso de autorización de la cuenta vía correo electrónico.
- **Configuración de Python en Visual Studio:** Este IDE tiene un sólido soporte para Python, incluyendo las versiones más recientes del lenguaje.

entorno de desarrollo completo con muchas características y ventajas de desarrollo.

- **Códigos de extracción:** Con la cuenta autorizada se crea código para la sustracción de comentarios en Twitter (tweets) de acuerdo a las palabras clave de filtrado.
- **Códigos de depuración:** Se genera código para realizar la actividad de depuración.
- **Uso de Técnica WebScraping:** Para los sitios Facebook y YouTube se utilizó esta estrategia para acceder a los sitios mencionados y descargar su contenido. Se hizo una extracción de los datos (comentarios) que se considera son de ayuda para el análisis y el objetivo deseado (siempre respetando las políticas de seguridad y privacidad estipulados).

4.1.4.2 Limpieza de datos

Aquí se consideran los datos que fueron recopilados y se observa que contienen ruido, como información irrelevante o duplicada. En esta etapa, se llevó a cabo un proceso de depuración de los datos con el objetivo de eliminar cualquier dato no deseado y garantizar la calidad y la integridad de los datos que serán utilizados en el análisis.

Para esta actividad fueron creados códigos en Python con el propósito de eliminar datos que no servían o no se consideran de utilidad. En la siguiente tabla se muestran las actividades de depuración aplicadas **FIGURA XXX**

Actividades de Depuración
Eliminar menciones y respuestas a tweets de usuarios, las cuales se identifican mediante cadenas que inician con @
Eliminar URL – Cadenas que inician con http://
Eliminar caracteres especiales como: #, &, "", {}, ", ' ,
Eliminar datos duplicados

Un ejemplo de un comentario en Twitter y el proceso de limpieza aplicado



Los #antivacunas deben pensar que sus seguidores son idiotas o no saben leer: la Prego publica en su canal esta noticia como muestra del peligro de las #vacunascovid cuando en la información se lee claramente que la fatiga del cantante se debe al #longcovid #SaludsinBulos

@MaarisscaVsBuks . Jun 20

Los antivacunas deben pensar que sus seguidores son idiotas o no saben leer: la Prego publica en su canal esta noticia como muestra del peligro de las #vacunascovid cuando en la información se lee claramente que la fatiga del cantante se debe al #longcovid #SaludsinBulos

4.1.4.3 Normalización de datos

Muchos de los datos recopilados contenían variaciones en el formato, como errores ortográficos, abreviaciones o diferentes formas de expresar conceptos similares. La normalización de datos consistió en estandarizar el formato y asegurarse de que los datos fueran coherentes y comparables.

4.1.4.4 Análisis Psicolingüístico

En este apartado el objetivo es identificar y extraer las características psicológicas y lingüísticas de los datos que han sido depurados. Estas características pueden ser utilizadas posteriormente en diversos análisis, como el análisis de sentimiento, la clasificación de textos, la detección de perfiles psicológicos, entre otros. Esta etapa es de suma importancia, ya que nos permite obtener vectores de características que serán utilizados posteriormente para entrenar los algoritmos de aprendizaje supervisado.

LIWC se utiliza en este proceso como analizador de textos para obtener las características psicolingüísticas y los posibles patrones de polaridad o sentimientos de acuerdo a las categorías de clasificación con las que cuenta la herramienta. La versión 15 que es utilizada en este trabajo de investigación cuenta con 86 categorías y éstas se encuentran clasificadas en diferentes dimensiones **figura XXX** (algunas dimensiones y categorías asociadas)

Psicológicas Básicas	Sociales	Cognitivas y Perceptuales	De Contenido
<ul style="list-style-type: none">• Afecto Positivo (<i>feliz, amor</i>)• Afecto Negativo (<i>triste, enojado</i>)• Cognición• Motivación	<ul style="list-style-type: none">• 1ra Persona Singular (<i>yo, mío</i>)• 3ra Persona (<i>él, ellos</i>)• Relaciones Sociales (<i>amigo, enemigo</i>)	<ul style="list-style-type: none">• Verbales• Perceptuales. Está relacionada con los sentidos (<i>ver, oír</i>)	<ul style="list-style-type: none">• Palabras Funcionales (<i>pronombres, preposiciones, conjunciones, etc.</i>)• Palabras Psicológicas (<i>estados mentales y emociones</i>)

Cada categoría contiene un conjunto de palabras relacionadas y se utiliza para medir la frecuencia de esas palabras en un texto dado. Las categorías se han desarrollado en base a investigaciones en psicología y lingüística. En este trabajo de investigación se utilizó la versión LIWC 2007 en español. **Figura XXX** (Sección con extracción de Categorías)

1	Funct	23	verbYO	43	MecCog	63	Espacio
2	TotPron	24	verbTU	44	Insight	64	Tiempo
3	PronPer	25	verbNOS	45	Causa	65	Trabajo
4	Yo	26	verbosEL	46	Discrep	66	Logro
5	Nosotros	27	verbELLOS	47	Tentat	67	Placer
6	TuUtd	28	Subjuntiv	48	Certeza	68	Hogar
7	ElElla	29	VosUtds	49	Inhib	69	Dinero
8	Ellos	30	formal	50	Incl	70	Relig
9	PronImp	31	informal	51	Excl	71	Muerte
10	Articulo	32	verbVos	52	Percept	72	Asentir
11	Verbos	33	Social	53	Ver	73	NoFluen
12	VerbAux	34	Familia	54	Oir	74	Relleno
13	Pasado	35	Amigos	55	Sentir	75	AllPunc
14	Present	36	Humanos	56	Biolog	76	Period
15	Futuro	37	Afect	57	Cuerpo	77	Comma
16	Adverb	38	EmoPos	58	Salud	78	QMark
17	Prepos	39	EmoNeg	59	Sexual	79	Exclam
18	Conjunc	40	Ansiedad	60	Ingerir	80	Apostro
19	Negacio	41	Enfado	61	Relativ	81	OtherP
20	Cuantif	42	Triste	62	Movim	82	Emoji
21	Numeros						
22	Maldec						

La clasificación de palabras en LIWC se basa en la comparación de cada palabra con un diccionario léxico predefinido que contiene miles de palabras y sus asignaciones de categorías. Cada palabra del texto analizado se compara con las palabras del diccionario y se asigna a las categorías correspondientes según las reglas establecidas, lo cual significa que pueden pertenecer a más de una categoría.

Por ejemplo, la palabra "abandon*" y sus diferentes derivaciones léxicas muestra las diferentes categorías en las que puede ser clasificada. Esto se puede observar en la

Figura XXX

ENTRY	Verbos	VerbAux	Pasado	Present	Futuro
abandon*					
abandona	X			X	
abandonamos	X		X	X	
abandonan	X			X	
abandonar					
abandonara*	X		X		
abandonaron	X		X		
abandonará*	X				X
abandonaré*	X				X
abandonas	X			X	
abandonase*	X		X		
abandonaste*	X		X		
abandone*	X			X	
abandono	X			X	
abandonáis	X			X	
abandonáramos	X		X		
abandonásemos	X		X		
abandoné	X		X		
	X			X	

Para el análisis de sentimientos LIWC utiliza un enfoque basado en el análisis de palabras para determinar el sentimiento o la carga emocional asociada a un texto. Aunque no es un enfoque de análisis de sentimientos, puede proporcionar una estimación general del tono emocional presente en un texto.

Utiliza categorías específicas relacionadas con emociones positivas y negativas para analizar el contenido emocional de un texto. Algunas de las categorías relevantes para el análisis de sentimientos en LIWC incluyen "Afecto positivo" y "Afecto negativo".

Durante el análisis, LIWC cuenta la frecuencia de las palabras que caen dentro de estas categorías emocionales en el texto. Por ejemplo, palabras como "feliz",

"alegre" y "amor" se asociarían con el afecto positivo, mientras que palabras como "triste", "enojado" y "odio" se asociarían con el afecto negativo.

También nos puede proporcionar información adicional sobre el análisis de sentimientos, como la intensidad emocional o la presencia de palabras relacionadas con emociones específicas. Sin embargo, es importante tener en cuenta que el análisis de sentimientos en LIWC se basa en palabras individuales y no tiene en cuenta el contexto completo de las oraciones o la coherencia semántica de las frases.

Es importante destacar que este análisis puede ser útil para obtener una visión general del tono emocional presente en un texto, pero puede no capturar todas las sutilezas y matices del sentimiento humano.

Con el propósito de realizar el procesamiento del texto utilizando LIWC, se procedió a convertir cada corpus en archivos de texto plano. Luego, se realizó un análisis individual de cada uno de los comentarios (Facebook y YouTube) y tweets. Tomemos en cuenta que LIWC procesa el texto en tres pasos:

- a) Detección de palabras presentes en su diccionario
- b) Asociación de las palabras del texto con las categorías definidas por LIWC
- c) Cálculo del porcentaje de palabras que se encuentran en cada categoría.

A continuación, se presenta un ejemplo de un tweet respecto al movimiento "antivacunas" y su proceso (de 3 pasos) a través de LIWC.

- a) **Detección de Palabras.** Durante el análisis del texto, LIWC busca coincidencias entre las palabras del texto y las palabras del diccionario. Se compara cada palabra del texto con las palabras del diccionario para determinar si hay coincidencias. En las Figuras XXX y XXX



En mi cuerpo mando yo. La etiqueta de antivacunas es injusta. El periodismo que ha obviado, ignorado, omitido otras voces igualmente científicas y médicas cuestionado la vacunación, ¿es también Negacionismo desde el otro lado? Tiempo al Tiempo.

Figura XXX

Word	Frequency
tiempo	2
injusta	1
cuestionado	1
vacunación	1
antivacunas	1
médicas	1
mando	1
vozes	1
igualmente	1
cuerpo	1
etiqueta	1
negacionismo	1
cientificas	1
obviado	1
ignorado	1
omitido	1
periodismo	1

- b) **Asociación de las palabras del texto con las categorías.** Si una palabra del texto coincide con una palabra del diccionario, se asigna a las categorías correspondientes definidas por LIWC. Esto implica que el texto recibirá una asignación a cada categoría correspondiente a las palabras encontradas en el diccionario. **Figura XXX Categorización de palabras**

	Artículo	Salud	Verbos	EmoNeg	MegCog	Prepos	Yo
En						X	
mi							X
cuerpo							
mando			X				
yo							X
La	X						
etiqueta							
de						X	
antivacunas		X		X	X		
es			X				
injusta							
El	X						
periodismo							
que							
ha						X	
obviado			X				
ignorado			X				
omitido			X				

d) Cálculo del porcentaje de palabras que se encuentran en cada categoría. Finalmente se construyen los vectores de características, que son las representaciones numéricas de los textos que se utilizan para describir y cuantificar la presencia de diferentes categorías lingüísticas y psicológicas en el texto analizado. Cada vector de características está compuesto por una serie de valores numéricos que representan la frecuencia o proporción de palabras pertenecientes a cada categoría en el texto.

Estos vectores se crean a partir del procesamiento del texto utilizando el diccionario léxico y las reglas de asignación de las categorías de LIWC. En cada texto se analiza y se cuentan las frecuencias, las cuales se representan como valores numéricos en el vector de características. Estos vectores proporcionan una visión general del contenido lingüístico y psicológico del texto. En la **Figura XXX** Se muestra una parte del esquema de vectorización del tweet analizado.

Filename	TotPron	PronPer	Yo	ElElla	PronImp	Articulo	Verbos	EmoNeg	MegCog	VerbAux	Present	Certeza	Inhib	Percept	Oir	Biolog	Salud
anti_20.txt	19.44	11.11	5.56	5.56	8.33	13.89	8.33	4.88	12.2	2.78	8.33	2.78	2.78	2.78	2.78	2.78	2.78

Los vectores de características generados a partir del análisis obtenido permiten cuantificar y representar numéricamente diferentes aspectos lingüísticos y psicológicos presentes en el texto. Estos vectores pueden utilizarse como entrada para algoritmos de aprendizaje automático o análisis estadístico, lo que permite realizar análisis cuantitativos y comparativos de textos en función de las categorías y dimensiones obtenidas en LIWC.

Para finalizar, se consideran todas las características que maneja LIWC, lo que resulta en la obtención de un vector de características único con una dimensión de 86 elementos. Estos vectores serán utilizados para el entrenamiento de algoritmos explicados a detalle en la siguiente sección.

3.1.5 Modelado

El objetivo principal de esta etapa es crear un modelo que pueda generalizar y hacer predicciones precisas sobre nuevos datos no vistos. Además, el proceso de modelado puede ayudar a descubrir relaciones ocultas, identificar variables importantes y proporcionar información útil para la toma de decisiones.

En esta etapa, se han seleccionado y aplicado algoritmos o modelos adecuados para realizar el análisis de los datos. Fueron ajustados los parámetros del modelo utilizando los datos de entrenamiento y se ha evaluado su rendimiento utilizando técnicas de validación y evaluación.

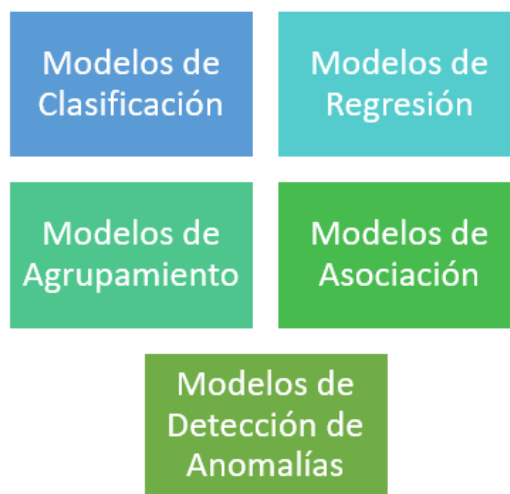
Para llevar a cabo esta tarea se ha utilizado la herramienta WEKA ya que nos proporciona una interfaz gráfica de usuario intuitiva y permite cargar conjuntos de datos, aplicar algoritmos de aprendizaje automático, evaluar y comparar modelos, y visualizar los resultados.

Algunas de las clasificaciones de modelos más comunes que se pueden encontrar en Weka son las siguientes:

- **Modelos de clasificación:** Son algoritmos y técnicas que se utilizan para clasificar instancias o datos en categorías o clases predefinidas.

- **Modelos de regresión:** Son algoritmos y técnicas utilizadas para predecir valores numéricos continuos en función de variables de entrada o características.
- **Modelos de agrupamiento:** Son algoritmos y técnicas utilizadas para agrupar instancias o datos en grupos o clústeres según su similitud o proximidad en el espacio de características. Estos modelos se utilizan para descubrir patrones o estructuras ocultas en los datos y agrupar instancias que comparten características similares.
- **Modelos de asociación:** Son algoritmos y técnicas utilizados para descubrir patrones interesantes de co-ocurrencia entre elementos en conjuntos de datos. Estos modelos se utilizan para encontrar relaciones entre diferentes elementos o atributos y revelar asociaciones frecuentes entre ellos.
- **Modelos de detección de anomalías:** Son algoritmos y técnicas utilizados para identificar patrones o instancias inusuales o anómalas en conjuntos de datos. Estos modelos se utilizan para encontrar valores atípicos o comportamientos inusuales que pueden ser indicativos de problemas, fraudes, errores o eventos inesperados.

Clasificación de Modelos en WEKA



Los algoritmos seleccionados para este trabajo de investigación fueron optados por las ventajas y efectividad que otorgan sus resultados en la toma de decisiones.

- **J48:** Es un algoritmo de aprendizaje automático basado en árboles de decisión. Es una implementación popular y ampliamente utilizada. La decisión de utilizarlo consistió en que:
 - Genera árboles de decisión que son fácilmente interpretables. Cada nodo en el árbol representa una característica o atributo, y las ramas representan las

diferentes opciones o valores posibles para ese atributo. Esto permite comprender cómo se toman las decisiones y qué características son las más influyentes.

- Puede manejar conjuntos de datos con atributos numéricos y categóricos. Es capaz de realizar divisiones en función de la información ganada, que es una medida de la importancia de un atributo en la clasificación. Esto hace que sea adecuado para una amplia gama de conjuntos de datos que contienen diferentes tipos de atributos.
 - Utiliza técnicas de poda para reducir la complejidad del árbol y evitar el sobreajuste. La poda ayuda a evitar que el árbol se ajuste demasiado a los datos de entrenamiento y mejora su capacidad de generalización a nuevos datos.
 - Es eficiente en términos de tiempo de ejecución y consumo de recursos. Puede manejar conjuntos de datos de tamaño considerable sin requerir una potencia de cómputo excesiva.
 - Hay una amplia documentación, ejemplos y recursos disponibles para ayudar en su implementación y comprensión.
- **BayesNet:** También conocido como Red Bayesiana, es un algoritmo de aprendizaje automático que utiliza redes bayesianas para modelar la relación probabilística entre las variables de un conjunto de datos. Las bondades que ofrece este modelo fueron consideradas para este trabajo ya que:
- Permite modelar las relaciones de dependencia probabilística entre las variables de un conjunto de datos. Esto puede ser especialmente útil cuando las variables tienen relaciones complejas y no lineales entre sí.
 - Las redes bayesianas proporcionan una representación gráfica intuitiva de las relaciones entre las variables. Esto permite comprender y visualizar cómo las diferentes variables influyen entre sí y cómo se propagan las probabilidades a través del modelo.
 - Son capaces de lidiar con la incertidumbre inherente en los datos y en las relaciones entre variables. Pueden manejar información incompleta o datos faltantes y realizar inferencias probabilísticas para estimar la probabilidad de diferentes resultados.
 - Puede aprender de manera incremental a medida que se le presentan nuevos datos. Puede adaptarse y actualizar el modelo a medida que se obtiene más información, lo que lo hace adecuado para entornos en los que los datos pueden cambiar o evolucionar con el tiempo.
 - Está diseñado para ser computacionalmente eficiente. Puede manejar conjuntos de datos de tamaño considerable sin requerir una gran cantidad de recursos computacionales.
- **SMO (Sequential Minimal Optimization):** Es un algoritmo de aprendizaje automático utilizado para construir modelos de clasificación basados en máquinas

de vectores de soporte (SVM). Fueron consideradas las ventajas de este modelo porque:

- Utiliza la técnica de máquinas de vectores de soporte, que busca encontrar el hiperplano óptimo que separa las clases en el espacio de características. Esto puede ser especialmente beneficioso cuando se trabaja con conjuntos de datos desafiantes o con características no lineales.
- Aprovecha técnicas de optimización y descomposición para dividir el problema en subproblemas más pequeños, lo que ayuda a reducir el tiempo de entrenamiento y el consumo de recursos.
- Permite la elección de diferentes funciones de kernel (función matemática que se utiliza para medir la similitud o distancia entre dos puntos en un espacio de características) para adaptarse a la naturaleza de los datos y las características del problema. Puedes seleccionar entre kernels lineales, polinomiales, radiales y sigmoiales, lo que te brinda la capacidad de modelar relaciones más complejas en los datos.
- Es una herramienta popular y bien establecida con una comunidad activa de usuarios y desarrolladores. Esto significa que hay una amplia documentación, recursos y soporte disponibles que me pueden ayudar a utilizar el modelo SMO y abordar cualquier problema o pregunta que pueda surgir en la utilización del mismo.

Para aplicar el aprendizaje automático hay que entrenar y evaluar los modelos elegidos (para las instancias obtenidas en LIWC); y para lograr esto se requiere tener dos conjuntos de datos:

- **Conjunto de entrenamiento:** Este subconjunto se utiliza para alimentar al modelo de aprendizaje automático durante su fase de entrenamiento. El modelo aprende a partir de las relaciones entre las características del conjunto de entrenamiento y las etiquetas correspondientes de la variable objetivo. El objetivo del entrenamiento es ajustar los parámetros internos del modelo para que pueda hacer predicciones precisas sobre datos no vistos.
- **Conjunto de prueba:** Una vez que el modelo ha sido entrenado, se evalúa su rendimiento utilizando el conjunto de prueba. Este subconjunto contiene ejemplos que el modelo no ha visto durante el entrenamiento. Se utiliza para medir la capacidad del modelo para generalizar y hacer predicciones precisas sobre datos nuevos y no vistos anteriormente. El conjunto de prueba proporciona una estimación imparcial del rendimiento real del modelo.

Una vez que se decide el modelo a trabajar, el algoritmo utiliza los datos proporcionados y crea un modelo en base a esta información; el cual puede

manifestarse en forma de reglas de clasificación, árboles de decisión o fórmulas matemáticas. Posteriormente, se realiza una validación del modelo con el conjunto de datos.

Para lograr esto se lleva a cabo una validación cruzada la cual consiste en dividir el conjunto de datos disponibles en múltiples subconjuntos y múltiples iteraciones de entrenamiento y así obtener una evaluación favorable del modelo.

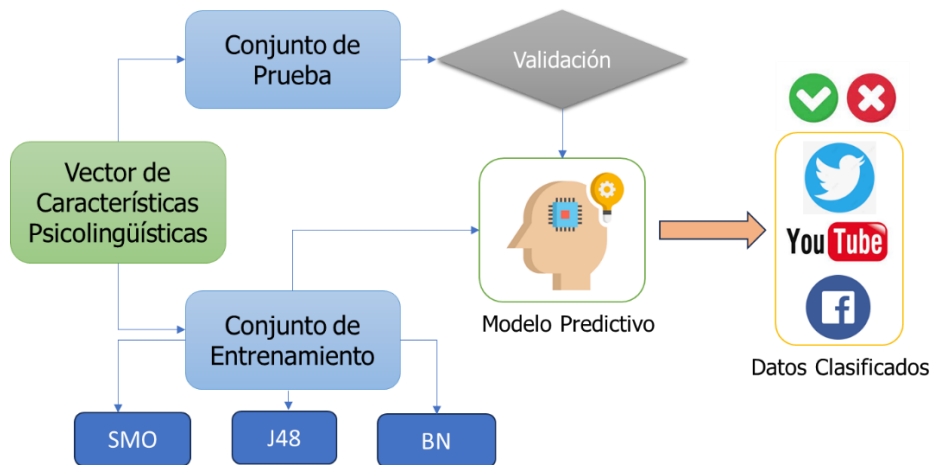
Dicha validación se llevó a cabo dividiendo el conjunto de datos en una muestra de 10 subconjuntos y a su vez el proceso crea una iteración de 10 repeticiones con el objetivo de utilizar un set de conjuntos para pruebas y otros para entrenamiento.

Validación Cruzada

Conjunto de Prueba		Conjunto de Entrenamiento	
		Subconjuntos	Iteraciones
Twitter	3000	300	10
Facebook	2000	200	10
YouTube	1500	150	10

Este proceso de validación permite obtener una evaluación más confiable y robusta del modelo, ya que se utiliza todo el conjunto de datos para entrenar y evaluar en diferentes combinaciones. También ayuda a detectar si el modelo tiene problemas de sobreajuste o subajuste, al evaluar su rendimiento en datos no vistos en cada iteración.

En la **Figura XXX** se muestra el procedimiento de aprendizaje automático para identificar los sentimientos de los movimientos antivacunas. Para lo cual fue necesario un conjunto de características representativas que se utilizan posteriormente para entrenar el clasificador y evaluar el modelo resultante. En este escenario, el modelo tiene la capacidad de diferenciar entre tweets que poseen contenido positivo y su relevancia y aquellos que no la contienen.



3.1.6 Evaluación

El objetivo principal es evaluar y validar los modelos de aprendizaje automático desarrollados, así como comparar diferentes enfoques y optimizar su rendimiento. Esto ayuda a determinar la efectividad y utilidad de los modelos en la solución del problema planteado y proporciona información para tomar decisiones y mejoras adicionales si es necesario.

Los resultados obtenidos en la evaluación del método para el análisis de sentimientos se presentan, analizan y evalúan con el objetivo de determinar la importancia de las características psicolingüísticas. Para valorar el rendimiento de estos modelos seleccionados fueron consideradas las siguientes métricas de evaluación:

- **Medida de Precisión:** Es una métrica utilizada en la evaluación de modelos de clasificación. Representa la proporción de instancias positivas correctamente clasificadas sobre todas las instancias clasificadas como positivas.

Es una medida útil cuando se quiere evaluar la capacidad del modelo para clasificar correctamente las instancias positivas, es decir, minimizar los falsos positivos. Es especialmente relevante cuando el costo de los falsos positivos es alto y se desea minimizar los errores de clasificación positiva. Algunas ventajas de utilizar la precisión como medida de evaluación son:

- La precisión se centra en la capacidad del modelo para identificar correctamente las instancias positivas, lo cual es importante en muchos escenarios donde los falsos positivos tienen consecuencias significativas.
- La precisión se expresa como un valor entre 0 y 1, donde 1 representa una alta precisión y 0 representa una baja precisión. Es fácil de interpretar y comprender su significado.
- La precisión puede combinarse con otras métricas, como la sensibilidad (recall) o la medida F1, para obtener una evaluación más completa del rendimiento del modelo. La fórmula aplicada a esta métrica se muestra en la **Figura XXX**

$$\text{Precisión} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Positivos}}$$

- **Medida de exhaustividad:** También conocida como recall o sensibilidad (Recall/Sensitivity), es una métrica utilizada en la evaluación de modelos de clasificación. Representa la proporción de instancias positivas correctamente clasificadas respecto al total de instancias que son realmente positivas.

Es una medida útil cuando se desea evaluar la capacidad del modelo para identificar correctamente todas las instancias positivas en un conjunto de datos. Es especialmente relevante cuando el costo de los falsos negativos es alto y se busca minimizar los errores de clasificación negativa. Las ventajas que pueden ser utilizadas de esta métrica son:

- Se centra en la capacidad del modelo para capturar correctamente las instancias positivas, lo cual es crucial en escenarios donde los falsos negativos pueden tener consecuencias graves.
- Se expresa como un valor entre 0 y 1, donde 1 representa una alta exhaustividad y 0 representa una baja exhaustividad. Es fácil de interpretar y comprender su significado.
- Es una métrica complementaria a la precisión, ya que la precisión se enfoca en los verdaderos positivos y la exhaustividad se enfoca en los falsos negativos. Juntas, estas métricas brindan una evaluación más completa del rendimiento del modelo. El cálculo de esta métrica se muestra en la **Figura XXX**

$$\text{Exhaustividad} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Negativos}}$$

- **La medida F:** También conocida como puntuación F o puntuación F1, es una métrica utilizada en la evaluación de modelos de clasificación que combina la precisión y la

exhaustividad en una sola medida. Su objetivo principal es encontrar un equilibrio entre estas dos métricas y proporcionar una evaluación más completa del rendimiento del modelo. Utilizar esta medida nos puede proporcionar ventajas como:

- Tiene en cuenta tanto los falsos positivos como los falsos negativos, brindando una evaluación equilibrada del rendimiento del modelo en términos de errores de clasificación tanto positivos como negativos.
- Al combinar la precisión y la exhaustividad en una sola medida, la puntuación F proporciona un equilibrio entre estas dos métricas. Esto es especialmente útil cuando se desea encontrar un compromiso óptimo entre minimizar los falsos positivos y los falsos negativos.
- Es especialmente útil cuando se trabaja con conjuntos de datos desequilibrados, es decir, cuando la proporción de instancias positivas y negativas es significativamente diferente. En tales casos, la precisión y la exhaustividad pueden ser engañosas por sí solas, y la medida F proporciona una evaluación más equilibrada y representativa del rendimiento del modelo. Para obtener una métrica F se realiza la operación que se muestra en la **Figura XXX**

$$F = 2 * \frac{\text{Precisión} * \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}}$$

Para completar la evaluación de los modelos es considerada la métrica Matriz de Confusión, la cual permite resumir el desempeño en términos de la clasificación correcta e incorrecta de las instancias. La matriz muestra la relación entre las clases reales y las clases predichas por el modelo.

Tiene una estructura cuadrada, donde las filas representan las clases reales y las columnas representan las clases predichas. Cada celda de la matriz muestra el número de instancias clasificadas correctamente e incorrectamente para una combinación específica de clase real y clase predicha. En la **Figura XXX** se puede observar la estructura de la matriz de confusión.

La matriz de confusión se utiliza para calcular varias métricas de evaluación, como **precisión, exhaustividad, puntuación F** y exactitud, entre otras. Al examinar los valores en la matriz de confusión, se pueden derivar las siguientes medidas:

Verdaderos positivos (True Positives, TP): Representa el número de instancias clasificadas correctamente como positivas.

Verdaderos negativos (True Negatives, TN): Representa el número de instancias clasificadas correctamente como negativas.

Falsos positivos (False Positives, FP): Representa el número de instancias clasificadas incorrectamente como positivas (clasificación errónea de negativas como positivas).

Falsos negativos (False Negatives, FN): Representa el número de instancias clasificadas incorrectamente como negativas (clasificación errónea de positivas como negativas).

Matriz de Confusión

		Predicción	
		Positivos	Negativos
Realidad	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

3.1.7 Despliegue

En esta última fase, se implementó el modelo. Esto puede implicar la integración de los resultados obtenidos que pueden ayudar a informar estrategias de comunicación, educación y divulgación que aborden las preocupaciones y mitos sobre las vacunas, promoviendo así una mayor aceptación y confianza en la inmunización.

CAPÍTULO IV RESULTADOS

En este apartado, presentamos los resultados obtenidos por los algoritmos Bayes Net, SMO, y J48 en términos de tres aspectos clave: análisis de sentimientos,

detección de temas principales y evaluación de la polaridad en relación con los movimientos antivacunas. Además, se examinaron los factores psicolingüísticos más prominentes asociados con el rechazo a las vacunas. Estos resultados nos brindan una visión detallada de las actitudes y emociones prevalentes en estos movimientos, así como de los factores subyacentes que influyen en la resistencia a la vacunación.

En cada uno de los modelos será utilizada una Clasificación Binaria en el cual se tienen únicamente dos clases o categorías posibles para cada instancia de datos. Es decir, el objetivo es asignar a cada ejemplo de entrada a una de las dos clases predefinidas. La idea es construir modelos o algoritmos que puedan aprender a distinguir y realizar predicciones precisas para cada una de las dos clases.

En esta investigación las clases que fueron clasificadas son: "provacunas" y "antivacunas", y para calcular las métricas será necesario tener un conjunto de datos etiquetados. Las predicciones realizadas por cada modelo a partir de esta información pueden ser calculadas de la siguiente forma:

Clasificación Binaria de las Clases: Provacunas y Antivacunas		
Verdaderos Positivos	VP	Son los casos en los que el modelo predice correctamente la clase "provacunas" y la etiqueta real también es "provacunas".
Falsos Positivos	FP	Son los casos en los que el modelo predice incorrectamente la clase "provacunas", pero la etiqueta real es "antivacunas".
Falsos Negativos	FN	Son los casos en los que el modelo predice incorrectamente la clase "antivacunas", pero la etiqueta real es "provacunas".

Las métricas de medición utilizadas son Precisión, Exhaustividad y F1 las cuales serán calculadas como se presenta a continuación:

Validación de Métricas		
Precisión	Se define la proporción de casos positivos predichos correctamente (VP) sobre el total de casos clasificados como positivos por el modelo	Medirá la exactitud de las predicciones positivas del modelo. Un valor alto de precisión indica que hay una alta proporción de predicciones positivas que son correctas.
Exhaustividad	Se define la proporción de casos positivos predichos correctamente (VP) sobre el total de casos reales que son positivos (VP + FN).	Medirá la capacidad del modelo para identificar todos los casos positivos en el conjunto de datos. Un valor alto de exhaustividad indica que el modelo es capaz de capturar la mayoría de los casos positivos.
F	Combina la precisión y la exhaustividad en una sola medida. Se calcula como la media armónica de ambas métricas.	Proporcionará un balance entre la precisión y la exhaustividad.

Veamos en las siguientes secciones con valores numéricos y probando varios modelos para calcular estas métricas.

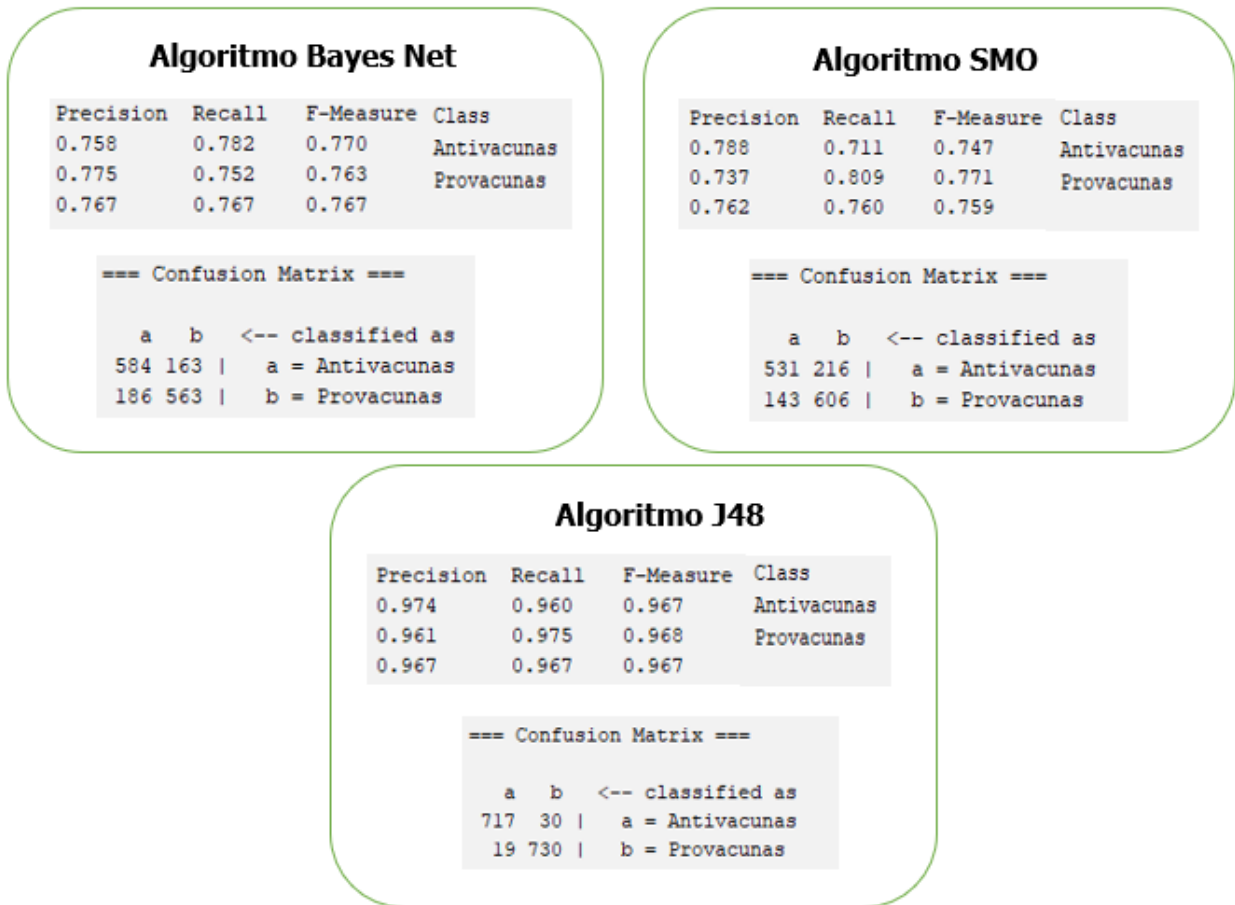
4.1 Resultados obtenidos en YouTube

A continuación, se muestran los resultados obtenidos para el dominio de YouTube en el cual se utilizó el corpus de referencia dividido en Provacunas y Antivacunas. Se presentarán los resultados más relevantes, acompañados de gráficos, tablas y análisis detallados para respaldar las conclusiones obtenidas. Se explorarán diferentes dimensiones y variables clave, ofreciendo una perspectiva completa y precisa de los resultados del análisis realizado en YouTube.

Lo que se pretende modelar y representar son las relaciones causales entre diferentes variables en un corpus establecido en el contexto de Provacunas y Antivacunas, esto puede ayudar a comprender cómo ciertos factores, creencias o argumentos influyen en la posición a favor o en contra de las vacunas. La estructura de la red puede capturar las dependencias entre las variables relevantes y proporcionar una representación clara de cómo estas relaciones interactúan.

Estos resultados pueden ayudar a estimar la probabilidad de que ciertos argumentos o afirmaciones estén relacionados con posturas Provacunas o Antivacunas. Esto puede ser útil para analizar la influencia de diferentes factores en las opiniones de

las personas sobre las vacunas. En la Figura XXX se pueden observar los resultados obtenidos de los diferentes Algoritmos aplicados al corpus de dominio para Youtube.



YouTube	Provacunas			Antivacunas		
	BayesNet	SMO	J48	BayesNet	SMO	J48
Precisión	0.775	0.737	0.961	0.758	0.788	0.974
Exhaustividad	0.752	0.809	0.975	0.782	0.711	0.960
F1	0.763	0.771	0.968	0.770	0.747	0.967

Cada valor de precisión obtenido en los tres modelos de clasificación para la categoría de "Provacunas" con resultados de 0.775 (77.5%) en el modelo Bayes Net, 0.737 (73.7%) en el algoritmo SMO y un valor de 0.961 (96.1%) para el modelo J48 significa que de los movimientos clasificados como "Provacunas" son realmente movimientos relacionados con esa categoría. Con respecto a la categoría "Antivacunas" los resultados obtenidos para el modelo Bayes Net es de 0.758

(75.8%), para el algoritmo SMO un valor de 0.788 (78.8%) y en el modelo J48 un valor de 0.974 (97.4%) lo cual nos da como respuesta que el porcentaje obtenido es la certeza que corresponde a cada categoría. Es decir, de todas las veces que el modelo predijo un movimiento como "Provacunas" o "Ativacunas", aproximadamente el porcentaje arrojado de esas predicciones fueron correctas. Los modelos tienen una tasa razonablemente alta de aciertos en las clasificaciones positivas.

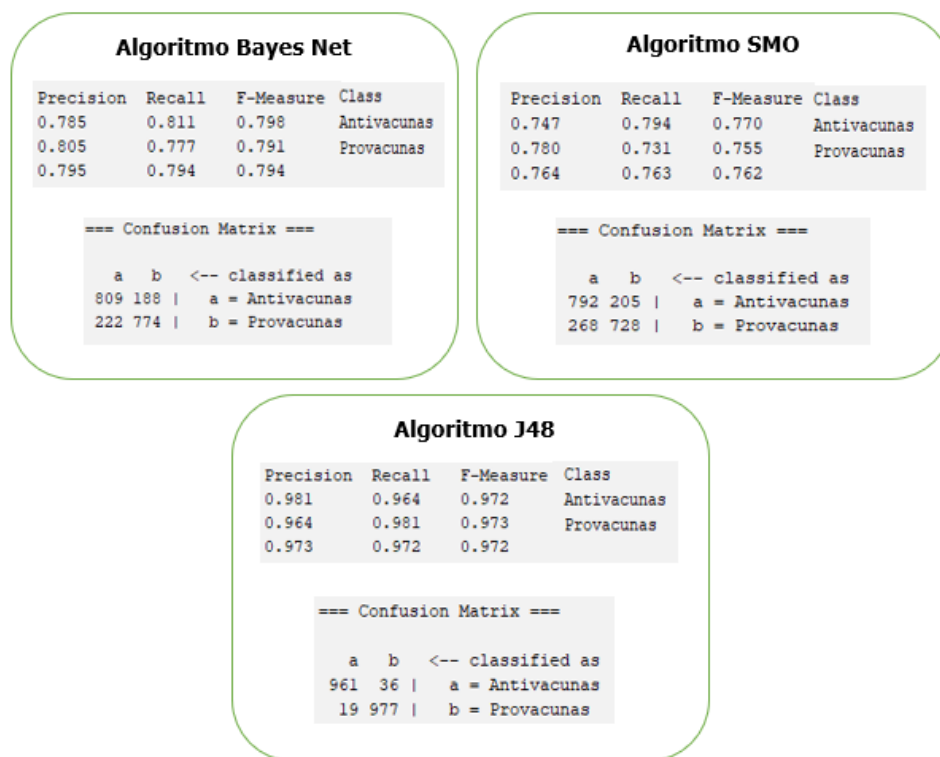
Los valores obtenidos en la métrica de exhaustividad por cada modelo en la clasificación de "Provacunas" nos muestra que para el algoritmo BayesNet tiene una valoración de 0.752 (75.2%), en el modelo SMO una estimación de 0.809 (80.9%) y para J48 un resultado del 0.975 (97.5%). En comparación con los valores obtenidos para la categoría "Antivacunas" lo arrojado por Bayes Net es un indicador del 0.782 (78.2%), el algoritmo SMO con un valor de 0.711 (71.1%) y para J48 el resultado obtenido es de 0.967 (96.7%). Esto significa que de todos los movimientos que en realidad son "Provacunas" o "Antivacunas", el modelo tiene una capacidad razonable para detectar y capturar los resultados positivos en el conjunto de datos.

Recordemos que F1 es una medida combinada de precisión y exhaustividad, por lo tanto, los resultados que arrojaron cada uno de los modelos para la categoría "Provacunas" respecto a esta métrica son en base a lo que resulta de cada uno de los valores antes mencionados y aplicados en cada algoritmo. En Bayes Net se obtiene un valor de 0.763 (76.3%), SMO muestra un resultado de 0.771 (77.1%) y J48 genera un sólido resultado de 0.968 (96.8%). Considerando la categoría "Antivacunas" los logros alcanzados por el modelo Bayes Net es el valor de 0.770 (77.0%), el algoritmo SMO generó un valor de 0.747 (74.7%) y el modelo J48 con un alcance de 0.967 (96.7%). Los valores obtenidos en esta métrica nos revelan el equilibrio y rendimiento razonablemente bueno entre la precisión y la exhaustividad de los modelos aplicados en la clasificación de las categorías "Provacunas" y "Antivacunas".

4.2 Resultados obtenidos en Facebook

En esta sección se exhiben los resultados obtenidos en el ámbito de Facebook al emplear el conjunto de datos de referencia dividido en "Provacunas" y "Antivacunas". Se mostrarán los resultados más destacados, respaldados por gráficos, tablas y un análisis detallado para respaldar las conclusiones alcanzadas. Se explorarán diversas dimensiones y variables importantes, brindando una perspectiva integral y precisa de los resultados del análisis realizado en Facebook.

La aplicación de modelos nos ayuda a identificar tendencias y patrones en los movimientos Provacunas y Antivacunas presentes en el corpus. Por ejemplo, puede revelar los temas más discutidos, las posturas predominantes o los argumentos comunes en cada categoría de los comentarios de Facebook. A menudo incluyen información contextual relevante, esta información puede ser útil para comprender mejor las opiniones y los factores que influyen en ellas, como las diferencias culturales o las experiencias personales. En la **Figura XXX** se presentan los resultados obtenidos al aplicar diferentes algoritmos al corpus específico de Facebook.



Facebook	Provacunas			Antivacunas		
	BayesNet	SMO	J48	BayesNet	SMO	J48
Precisión	0.805	0.780	0.964	0.785	0.747	0.981
Exhaustividad	0.777	0.731	0.981	0.811	0.794	0.964
F1	0.791	0.755	0.973	0.798	0.770	0.972

Cada modelo de clasificación obtuvo los siguientes valores de precisión para la categoría "Provacunas": 0.805 (80.5%) en Bayes Net, 0.780 (78.0%) en SMO y 0.964 (96.4%) en J48. Estos valores indican la proporción de movimientos clasificados correctamente como "Provacunas" en relación con todos los movimientos clasificados como tal. Para la categoría "Antivacunas", los resultados son los siguientes: 0.785 (78.5%) en Bayes Net, 0.747 (74.7%) en SMO y 0.981 (98.1%) en J48. Estos valores representan la certeza de que los movimientos clasificados como "Antivacunas" son realmente movimientos relacionados con esa categoría. Los modelos muestran una tasa razonablemente alta de aciertos en las clasificaciones positivas.

En cuanto a la métrica de exhaustividad, los valores obtenidos para la categoría "Provacunas" son: 0.777 (77.7%) en Bayes Net, 0.731 (73.1%) en SMO y 0.981 (98.1%) en J48. Para la categoría "Antivacunas", los valores son: 0.811 (81.1%) en Bayes Net, 0.794 (79.4%) en SMO y 0.964 (96.4%) en J48. Estos resultados indican la capacidad del modelo para detectar y capturar los resultados positivos en el conjunto de datos.

La métrica F1, muestra los siguientes resultados para la categoría "Provacunas": 0.791 (79.1%) en Bayes Net, 0.755 (75.5%) en SMO y 0.973 (97.3%) en J48. En la categoría "Antivacunas", los valores son: 0.798 (79.8%) en Bayes Net, 0.770 (77.0%) en SMO y 0.972 (97.2%) en J48. Estos valores reflejan el equilibrio y el rendimiento razonablemente bueno entre la precisión y la exhaustividad de los modelos aplicados en la clasificación de las categorías "Provacunas" y "Antivacunas".

4.3 Resultados obtenidos en Twitter

En este último grupo se muestran los resultados obtenidos para el dominio de Twitter en el cual se utilizó el corpus de referencia dividido en Provacunas y Antivacunas. Se presentarán los resultados más relevantes, acompañados de gráficos, tablas y análisis detallados para respaldar las conclusiones obtenidas. Se explorarán diferentes dimensiones y variables clave, ofreciendo una perspectiva completa y precisa de los resultados del análisis realizado en Twitter.

El objetivo de trabajar cada uno de estos modelos es que al tener una amplia gama de opiniones y perspectivas sobre el tema de las vacunas. La gran diversidad de culturas independientes al idioma español enriquece el conjunto de datos y brinda una visión más completa del tema que se está analizando. Estos datos son de gran utilidad para esta investigación ya que nos pueden ayudar en el análisis de sentimientos, identificación de tendencias y una buena toma de decisiones. En la **Figura XXX** se muestran los resultados obtenidos al aplicar los distintos algoritmos seleccionados al corpus de Twitter.

Algoritmo Bayes Net

Precision	Recall	F-Measure	Class
0.860	0.909	0.884	Antivacunas
0.903	0.852	0.877	Provacunas
0.882	0.880	0.880	

```

=== Confusion Matrix ===
      a  b  <-- classified as
1361 137 |  a = Antivacunas
 221 1276 |  b = Provacunas
    
```

Algoritmo SMO

Precision	Recall	F-Measure	Class
0.831	0.966	0.893	Antivacunas
0.959	0.803	0.874	Provacunas
0.895	0.884	0.884	

```

=== Confusion Matrix ===
      a  b  <-- classified as
1447  51 |  a = Antivacunas
 295 1202 |  b = Provacunas
    
```

Algoritmo J48

Precision	Recall	F-Measure	Class
0.980	0.960	0.970	Antivacunas
0.961	0.981	0.971	Provacunas
0.970	0.970	0.970	

```

=== Confusion Matrix ===
      a  b  <-- classified as
1438  60 |  a = Antivacunas
  29 1468 |  b = Provacunas
    
```

Twitter	Provacunas			Antivacunas		
	BayesNet	SMO	j48	BayesNet	SMO	J48
Precisión	0.903	0.959	0.961	0.860	0.831	0.980
Exhaustividad	0.852	0.803	0.981	0.909	0.966	0.960
F1	0.877	0.874	0.971	0.884	0.893	0.970

Los resultados obtenidos en los tres modelos de clasificación para la categoría "Provacunas" fueron los siguientes: el modelo Bayes Net obtuvo una precisión del 0.903 (90.3%), el algoritmo SMO obtuvo una precisión del 0.959 (95.9%), y el modelo J48 obtuvo una precisión del 0.961 (96.1%). Estos valores indican que, en promedio, el porcentaje de movimientos clasificados correctamente como "Provacunas" en cada modelo es el mencionado. Lo mismo sucede para la categoría "Antivacunas", donde el modelo Bayes Net obtuvo una precisión del 0.860 (86.0%), el algoritmo SMO una precisión del 0.831 (83.1%), y el modelo J48 una precisión

del 0.980 (98.0%). Estos porcentajes representan la certeza de que los movimientos clasificados como "Provacunas" o "Antivacunas" son realmente movimientos relacionados con esas categorías. Los modelos tienen una tasa razonablemente alta de aciertos en las clasificaciones positivas.

En cuanto a la métrica de exhaustividad, el algoritmo Bayes Net obtuvo un valor del 0.852 (85.2%) en la clasificación de "Provacunas", el modelo SMO obtuvo un valor del 0.803 (80.3%), y el modelo J48 obtuvo un valor del 0.981 (98.1%). Para la categoría "Antivacunas", el algoritmo Bayes Net obtuvo un valor del 0.909 (90.9%), el algoritmo SMO un valor del 0.966 (96.6%), y el modelo J48 un valor del 0.960 (96.0%). Esto significa que los modelos tienen una capacidad razonable para detectar y capturar correctamente los movimientos positivos en el conjunto de datos, tanto en la categoría "Provacunas" como en la categoría "Antivacunas".

Resumiendo sabemos que el valor de F1 es una medida combinada de precisión y exhaustividad. En el caso de la categoría "Provacunas", los modelos arrojaron los siguientes valores de F1: Bayes Net (0.877, es decir, 87.7%), SMO (0.874, es decir, 87.4%) y J48 (0.971, es decir, 97.1%). Para la categoría "Antivacunas", los valores de F1 fueron: Bayes Net (0.884, es decir, 88.4%), SMO (0.893, es decir, 89.3%) y J48 (0.970, es decir, 97.0%). Estos valores revelan un equilibrio y un rendimiento razonablemente bueno entre la precisión y la exhaustividad de los modelos aplicados en la clasificación de las categorías "Provacunas" y "Antivacunas".

4.4 Análisis y Discusión de los Resultados

En esta sección, presentamos los resultados obtenidos de nuestro análisis en profundidad de los comentarios relacionados con el tema de las vacunas. Nuestro estudio se centró en la clasificación de los comentarios en dos categorías principales: "Provacunas" y "Antivacunas".

A través de la aplicación de diferentes algoritmos de clasificación, evaluamos la precisión, exhaustividad y puntuación F1 de cada modelo. En la **tabla XXX** se pueden observar los resultados obtenidos por cada uno de los Modelos.

Corpus	Modelo	Medio	Métrica		
			Precision	Exhaustividad	F1
Provacunas	J48	Twitter	0.961	0.981	0.971
		Facebook	0.964	0.981	0.973
		YouTube	0.961	0.975	0.968
	Bayes Net	Twitter	0.903	0.852	0.877
		Facebook	0.805	0.777	0.791
		YouTube	0.775	0.752	0.763
	SMO	Twitter	0.959	0.803	0.874
		Facebook	0.780	0.731	0.755
		YouTube	0.737	0.809	0.771
Antivacunas	J48	Twitter	0.980	0.960	0.970
		Facebook	0.981	0.964	0.972
		YouTube	0.974	0.960	0.967
	Bayes Net	Twitter	0.860	0.909	0.884
		Facebook	0.785	0.811	0.798
		YouTube	0.758	0.782	0.770
	SMO	Twitter	0.831	0.966	0.893
		Facebook	0.747	0.794	0.770
		YouTube	0.788	0.711	0.747

Los resultados del estudio indicaron que el algoritmo J48 mostró los mejores resultados en términos de precisión y rendimiento. Demostró un alto porcentaje de aciertos en la clasificación de los datos, superando a los otros algoritmos evaluados. Sus resultados se destacaron por su precisión y capacidad para manejar diferentes tipos de datos.

Se observa que el algoritmo J48 mostró los mejores resultados en términos de precisión en la clasificación de datos en las redes sociales. Específicamente, Twitter obtuvo los porcentajes más altos de aciertos utilizando este algoritmo, seguido de Facebook y YouTube en orden descendente.

En cuanto al algoritmo Bayes Net, se encontró que obtuvo resultados medios en comparación con J48. En Twitter, también logró los porcentajes más altos, seguido de Facebook y YouTube.

Por otro lado, los resultados del algoritmo SMO fueron los más bajos en los tres medios de redes sociales evaluados, con los porcentajes más altos en Twitter, seguido de Facebook y YouTube.

Aunque los resultados de Bayes Net y SMO fueron inferiores en comparación con J48, es importante destacar que estos algoritmos lograron resultados por encima de la media establecida. Esto indica que, si bien no alcanzaron el nivel de rendimiento de J48, aún demostraron ser resultados muy buenos y eficaces en la clasificación de datos en las plataformas de redes sociales.

Se puede concluir que el algoritmo J48 obtuvo los mejores resultados, con Twitter liderando en porcentajes de aciertos. Bayes Net y SMO obtuvieron resultados medios y bajos respectivamente, pero aún se destacaron al superar la media establecida. Aunque los resultados de J48 fueron superiores, los resultados obtenidos por Bayes Net y SMO son considerados muy buenos en general.

A continuación, se presentan las características más relevantes o informativas de nuestro conjunto de datos y que contribuyen de manera significativa al rendimiento de cada uno de los modelos y se muestra un análisis psicolingüístico, psicológico y cognitivo de cada una de ellas. **La Figura XXX** muestra las características con mayor peso seleccionadas utilizando el método de ganancia de la información.

Pronombres Personales	Negación	verbos	Familia	Emociones Positivas	Emociones Negativas	Ansiedad	Enfado	Cuerpo	Salud	Muerte	Period	QMark	Exclam
-----------------------	----------	--------	---------	---------------------	---------------------	----------	--------	--------	-------	--------	--------	-------	--------

4.4.1 Pronombres Personales

Los pronombres personales son palabras que reemplazan a los sustantivos o nombres en una oración. Estas palabras se utilizan para referirse a personas, animales u objetos sin necesidad de repetirlos explícitamente (UPB, 2020). Los pronombres personales cumplen la función de sustituir al sujeto o a otras entidades mencionadas previamente en el discurso, permitiendo una comunicación más fluida y evitando repeticiones innecesarias. En un contexto relacionado con las vacunas

éstos pueden ser analizados de forma Psicolingüística, Psicológica y de forma Cognitiva.

4.4.1.1 Análisis Psicolingüístico

- Puede indicar un enfoque en la experiencia y perspectiva personal de las personas que emitieron sus comentarios. Ellos pueden estar expresando sus propias creencias, experiencias y opiniones relacionadas con las vacunas, lo que refleja un sentido de identidad y autoexpresión (***Autoenfoque***).
- Puede reflejar una comunicación más íntima y directa. La gente puede estar buscando establecer una conexión personal con los lectores o destinatarios del mensaje, compartiendo sus puntos de vista y experiencias en primera persona (***Personalización del discurso***).
- Puede revelar una identificación y sentido de pertenencia a un grupo o comunidad específica. Los que comentan en un medio específico pueden estar comunicando su alineación con otros individuos que comparten sus opiniones sobre las vacunas (***Identificación y pertenencia***).

4.4.1.2 Análisis Psicológico

- Puede reflejar una necesidad de afirmar la propia identidad y perspectiva en el tema de las vacunas (***Autoafirmación y expresión de identidad***).
- Puede mostrar un sentido de apego emocional y valorativo a las creencias y posturas sobre las vacunas. Se mantiene la postura de estar comunicando la importancia personal y la adhesión a sus convicciones en el tema (***Apego a creencias y valores***).
- Puede manifestar una intención de establecer una conexión emocional con otras personas o lectores. Los pronombres personales pueden ser utilizados para compartir experiencias personales, generar empatía y buscar apoyo o validación en un contexto de discusión sobre las vacunas (***Empatía y conexión emocional***).
- Puede resaltar la responsabilidad individual en la toma de decisiones relacionadas con las vacunas. Las personas pueden estar enfocándose en su propia libertad de elección y autonomía en el tema, enfatizando la importancia de la toma de decisiones personales (***Enfoque en la responsabilidad individual***).

4.4.1.3 Análisis Cognitivo

- Refleja un enfoque cognitivo (adquisición de conocimiento) centrado en la experiencia individual. El individuo puede estar priorizando su propia percepción y vivencias en relación con las vacunas (***Enfoque en la experiencia personal***).
- Muestra una identificación subjetiva con el contenido de los comentarios. Las personas pueden estar expresando su perspectiva personal y cómo las vacunas

impactan en su propia vida, lo que sugiere una conexión cognitiva y emocional con el tema (*Identificación subjetiva*).

- Cada una de los individuos pueden estar buscando validar sus propias creencias y puntos de vista, reforzando su sentido de identidad y seguridad cognitiva en relación con las vacunas (*Autoafirmación y autorreferencia*).

4.4.2 Negación

La negación se refiere a la acción de contradecir o rechazar una afirmación, indicando que algo carece de existencia o veracidad. La negación se puede definir como un mecanismo de defensa utilizado por las personas para hacer frente a un conflicto, negando su existencia o minimizando su impacto en su bienestar personal (Pérez Porto, J., Gardey, 2023).

4.4.2.1 Análisis Psicolingüístico

- Puede indicar una actitud de duda o escepticismo hacia las vacunas. Los individuos pueden estar expresando su desacuerdo o incredulidad respecto a la efectividad, seguridad u otros aspectos relacionados con las vacunas (*Expresión de dudas o escepticismo*).
- Puede reflejar una respuesta directa a argumentos o afirmaciones contrarias a las propias creencias sobre las vacunas. Las personas pueden utilizar la negación para refutar o desacreditar puntos de vista opuestos (*Contrarrestar argumentos contrarios*).
- Puede revelar una expresión de desacuerdo o rechazo hacia ciertas ideas o afirmaciones relacionadas con las vacunas. La gente puede utilizar la negación como una forma de expresar su oposición o discrepancia (*Manifestación de desacuerdo o disconformidad*).

4.4.2.2 Análisis Psicológico

- Puede reflejar una resistencia psicológica hacia la aceptación de nuevas ideas o información sobre las vacunas. Las personas pueden sentir miedo, incertidumbre o ansiedad ante la posibilidad de modificar sus creencias existentes, lo que puede llevar a la negación como una forma de protección psicológica (*Resistencia al cambio*).
- Puede actuar como un mecanismo de defensa psicológica para proteger la autoimagen y el bienestar emocional. Las personas pueden negar ciertos aspectos relacionados con las vacunas para preservar una sensación de control, seguridad o coherencia con sus creencias y valores fundamentales (*Mecanismo de defensa*).

- Puede ser una forma de mantener la coherencia cognitiva. Las personas pueden negar ciertos aspectos relacionados con las vacunas para evitar conflictos internos y mantener una visión coherente de sí mismas y del mundo que las rodea (***Necesidad de coherencia cognitiva***).

4.4.2.3 Análisis Cognitivo

- Puede reflejar una evaluación cognitiva crítica de la información relacionada con las vacunas. Los individuos pueden estar cuestionando o rechazando afirmaciones o datos que consideran no confiables, contradictorios o poco convincentes (***Diferencias en la valoración de la información***).
- Puede implicar un proceso cognitivo de sopesar los riesgos y beneficios asociados con las vacunas. La gente puede estar expresando preocupación o desconfianza en relación con los posibles efectos adversos o la efectividad de las vacunas, lo que influye en su proceso de toma de decisiones (***Evaluación de riesgos y beneficios***).
- Puede estar relacionado con la protección y defensa de creencias previas arraigadas sobre las vacunas. Los seres humanos pueden utilizar negaciones para mantener la coherencia cognitiva y proteger sus convicciones o valores fundamentales relacionados con la salud, la libertad individual o la autonomía (***Defensa de creencias previas***).

4.4.3 Verbos

Los verbos son términos lingüísticos que transmiten una acción, existencia o estado dentro de una oración. Son las palabras encargadas de indicar lo que el sujeto está realizando y desempeñan un papel fundamental como elemento central de la oración. De hecho, es imposible construir un pensamiento o una oración completa sin la inclusión de un verbo (Artmann, 2020).

4.4.3.1 Análisis Psicolingüístico

- Puede indicar una orientación hacia la expresión de acciones y opiniones relacionadas con las vacunas. Los individuos pueden estar describiendo sus propias experiencias, recomendando ciertas acciones o expresando sus puntos de vista de manera activa y dinámica (***Expresión de acciones y opiniones***).
- Puede reflejar un compromiso y participación activa en la discusión sobre las vacunas. Los seres humanos pueden estar involucrados emocionalmente en el tema y utilizar verbos para transmitir su participación activa en la conversación (***Manifestación de compromiso y participación***).

- Puede sugerir un enfoque en la acción y el cambio en el discurso sobre las vacunas. Las personas pueden estar enfocados en discutir medidas concretas, iniciativas o acciones que consideran necesarias para abordar la problemática de las vacunas (**Enfoque en la acción y el cambio**).

4.4.3.2 Análisis Psicológico

- Puede reflejar una tendencia hacia la acción y la toma de decisiones en relación con las vacunas. Los seres humanos pueden tener una motivación para expresar su compromiso y participación activa en el tema, reflejando una mentalidad orientada a la acción y la búsqueda de soluciones (**Orientación hacia la acción**).
- Puede ser una forma de expresar actitudes y emociones relacionadas con las vacunas. Los individuos pueden utilizar verbos para transmitir su entusiasmo, preocupación, esperanza, temor u otras emociones asociadas con las vacunas y su impacto en la salud (**Manifestación de actitudes y emociones**).
- Puede manifestar un sentido de control y empoderamiento en relación con las decisiones sobre las vacunas. Las personas pueden utilizar verbos para comunicar su capacidad de tomar decisiones informadas, actuar de acuerdo con sus creencias y tener un papel activo en el cuidado de su salud y la de los demás (**Expresión de control y empoderamiento**).

4.4.3.3 Análisis Cognitivo

- Puede reflejar un procesamiento cognitivo activo en relación con las vacunas. Los individuos pueden estar involucrados en el análisis y la discusión de información, evaluando opciones, generando ideas y tomando decisiones relacionadas con las vacunas (**Procesamiento activo de la información**).
- Puede indicar un enfoque cognitivo en la acción y los resultados con respecto a las vacunas. Las personas pueden estar interesados en discutir las implicaciones prácticas de las vacunas, cómo se implementan, los efectos esperados y los resultados que se pueden lograr (**Enfoque en la acción y los resultados**).
- puede mostrar un proceso cognitivo de evaluar las causas y consecuencias de las decisiones relacionadas con las vacunas. Los seres humanos pueden analizar las implicaciones de diferentes opciones y considerar las ramificaciones a corto y largo plazo. (**Evaluación de causas y consecuencias**)

4.4.4 Familia

Se utiliza el concepto de "familia" para referirse a un conjunto de individuos que comparten un vínculo de parentesco y conviven juntos en una unidad. Desde la

perspectiva sociológica, la noción de familia se refiere a la estructura social básica conformada por el padre, la madre y sus hijos (Significados, 2019).

4.4.4.1 Análisis Psicolingüístico

- Puede sugerir la importancia que se le atribuye al papel de la familia en el contexto de las vacunas. Los individuos pueden estar expresando la relevancia de las opiniones, creencias y decisiones de la familia en la elección de vacunarse o no (***Valoración de la influencia familiar***).
- Puede indicar la conexión emocional y las experiencias compartidas en torno a las vacunas. Las personas pueden estar compartiendo anécdotas, testimonios o reflexiones personales sobre cómo las decisiones de la familia han afectado su percepción o experiencia con las vacunas (***Experiencias compartidas***).
- Puede reflejar la consideración de diferentes generaciones en el tema de las vacunas. Las personas pueden estar teniendo en cuenta la opinión de padres, abuelos u otros miembros de la familia y cómo esto influye en sus propias creencias o decisiones (***Perspectiva intergeneracional***).
- Puede señalar tanto el apoyo como la resistencia familiar hacia las vacunas. La gente puede compartir experiencias positivas en las que la familia respalda la vacunación, o bien, relatar situaciones en las que la familia se opone a las vacunas y cómo esto ha influido en su perspectiva (***Apoyo y resistencia familiar***).

4.4.4.2 Análisis Psicológico

- Puede reflejar la importancia de las relaciones familiares en la formación de creencias y actitudes hacia las vacunas. Las personas pueden estar expresando el impacto emocional y social que la familia tiene en su decisión de vacunarse o no (***Influencia social y emocional***).
- Puede indicar un sentido de identificación y pertenencia a un grupo familiar en el contexto de las vacunas. Los individuos pueden sentirse vinculados con los valores, tradiciones y normas de su familia, lo que influye en su postura y decisiones relacionadas con las vacunas (***Identificación y pertenencia***).
- Puede sugerir la importancia de proteger a los miembros de la familia a través de decisiones relacionadas con las vacunas. La gente puede estar motivada por el deseo de garantizar la salud y el bienestar de sus seres queridos, lo que influye en su posición sobre la vacunación (***Responsabilidad y protección***).
- Puede señalar la existencia de conflictos y diferencias de opinión dentro de la familia en relación con las vacunas. Los seres humanos comparten sus experiencias de desacuerdo o resistencia por parte de algunos miembros de la familia, lo que genera tensiones y dilemas en su toma de decisiones (***Conflicto y diferencias de opinión***).

4.4.4.3 Análisis Cognitivo

- Puede reflejar la presencia de asociaciones cognitivas y conexiones personales entre las vacunas y la esfera familiar. Las personas pueden estar relacionando las decisiones y creencias sobre las vacunas con su experiencia y conocimiento en el ámbito familiar (***Asociaciones cognitivas y conexiones personales***).
- Puede indicar la consideración de diferentes perspectivas y puntos de vista relacionados con las vacunas dentro del contexto familiar. Los individuos pueden estar reflexionando sobre cómo las creencias y opiniones de los miembros de la familia pueden influir en sus propias percepciones y decisiones (***Consideración de perspectivas múltiples***).
- Puede sugerir la evaluación cognitiva de los beneficios y riesgos asociados con las vacunas en relación con el bienestar familiar. La gente puede estar sopesando los potenciales beneficios para la salud de la familia frente a posibles riesgos o preocupaciones (***Evaluación de beneficios y riesgos***).
- Puede mostrar la influencia cognitiva de la construcción de identidad y los roles familiares en las actitudes hacia las vacunas. Los seres humanos pueden estar considerando cómo su papel en la familia, como padre, madre, hijo/a u otro miembro, influye en su percepción y toma de decisiones en relación con las vacunas (***Construcción de identidad y roles familiares***).

4.4.5 Emociones Positivas

Las emociones positivas se refieren a aquellas que generan una sensación de bienestar, es decir, una experiencia emocional agradable, satisfactoria y deseada. Generalmente surgen como resultado de situaciones que percibimos como favorables y su duración suele ser breve. Estas emociones tienen un impacto directo en el bienestar y la salud de las personas, ya que experimentarlas ayuda a disminuir los niveles de tristeza, estrés y ansiedad (AutoGenosis, 2023).

4.4.5.1 Análisis Psicolingüístico

- Puede reflejar una actitud favorable hacia las vacunas. Las personas pueden expresar sentimientos de entusiasmo, satisfacción, confianza o esperanza relacionados con la efectividad de las vacunas y los beneficios que brindan (***Actitud favorable***).
- Puede servir como una forma de validación emocional para aquellos que apoyan las vacunas. Los individuos pueden utilizar palabras y expresiones positivas para

reforzar su apoyo y fomentar una respuesta emocional positiva en otros lectores **(Validación emocional)**.

- Puede indicar la voluntad de compartir experiencias positivas relacionadas con las vacunas. La gente puede relatar historias de éxito personal, testimonios de mejoría de salud o relatos de eventos felices que están relacionados con las vacunas **(Compartir experiencias positivas)**.
- Puede ayudar a construir confianza y optimismo en torno a las vacunas. Los seres humanos suelen utilizar palabras y expresiones que generen un ambiente de confianza y seguridad, transmitiendo la idea de que las vacunas son una solución efectiva y prometedora **(Construcción de confianza y optimismo)**.

4.4.5.2 Análisis Psicológico

- Puede reflejar una actitud favorable y una sensación de satisfacción en relación con las vacunas. Los individuos pueden experimentar emociones como alegría, gratitud, alivio o esperanza al hablar sobre las vacunas y los beneficios que perciben **(Actitud favorable y satisfacción)**.
- Puede servir como una forma de validar y apoyar emocionalmente a aquellos que tienen una postura favorable hacia las vacunas. Las personas pueden expresar solidaridad, empatía y respaldo hacia otros individuos que comparten su visión y experiencia positiva con las vacunas **(Validación y apoyo emocional)**.
- Puede indicar que la gente ha tenido experiencias positivas personales con las vacunas. Pueden estar compartiendo testimonios y narrativas relacionadas con mejoras en la salud, prevención de enfermedades o eventos felices vinculados a las vacunas Experiencias positivas personales **(Consideración de perspectivas múltiples)**.
- Puede reforzar las creencias y actitudes positivas hacia las vacunas. Los que comentan pueden buscar fortalecer su propia perspectiva y la de otros al enfatizar los aspectos positivos y prometedores de las vacunas, generando un sentido de confianza y seguridad **(Reforzamiento de creencias y actitudes)**.

4.4.5.3 Análisis Cognitivo

- Puede reflejar una evaluación cognitiva positiva de las vacunas. Los individuos pueden tener creencias y percepciones positivas sobre la eficacia, seguridad y beneficios de las vacunas, lo que se refleja en las emociones positivas expresadas **(Evaluación favorable de las vacunas)**.
- Puede indicar una atención selectiva a los aspectos positivos relacionados con las vacunas. Las personas pueden enfocarse en información, experiencias y testimonios que respalden su visión positiva, lo que influye en las emociones positivas asociadas con las vacunas **(Atención selectiva a aspectos positivos)**.

- Puede revelar creencias cognitivas optimistas y esperanzadoras sobre las vacunas. La gente puede tener confianza en el potencial de las vacunas para prevenir enfermedades, mejorar la salud y generar un impacto positivo en la sociedad, lo que genera emociones positivas relacionadas (**Creencias optimistas y esperanzadoras**).
- Puede reflejar un sesgo cognitivo hacia lo positivo en la interpretación de la información relacionada con las vacunas. Los seres humanos tienden a enfocarse en aspectos favorables y descartar o minimizar información negativa, lo que se refleja en las emociones positivas predominantes (**Sesgo cognitivo hacia lo positivo**).

4.4.6 Emociones Negativas

Una emoción negativa se refiere a una sensación desagradable o un sentimiento negativo que experimentamos, pero que en realidad cumple un papel importante en nuestra supervivencia y seguridad. Estas emociones actúan como señales de advertencia, y si se ignoran de manera constante, pueden dar lugar a problemas graves tanto a nivel individual como en la vida en general (IEPP - Instituto Europeo de Emociones Positivas., 2023).

4.4.6.1 Análisis Psicolingüístico

- Puede reflejar actitudes desfavorables hacia las vacunas. Los individuos pueden expresar sentimientos de preocupación, miedo, desconfianza o disgusto en relación con las vacunas y los posibles riesgos que perciben (**Actitudes desfavorables**).
- Puede servir como una forma de expresar frustración o descontento con aspectos relacionados con las vacunas. La gente puede manifestar insatisfacción con la efectividad, la seguridad, la distribución o las políticas de vacunación, lo que se refleja en las emociones negativas expresadas (**Expresión de frustración o descontento**).
- Puede indicar la validación emocional de experiencias negativas relacionadas con las vacunas. Los que comentan pueden compartir testimonios de eventos adversos, efectos secundarios o experiencias negativas vividas o percibidas, lo que genera emociones negativas compartidas (**Validación emocional de experiencias negativas**).
- Puede sugerir la expresión de preocupaciones o dudas sobre las vacunas. Los seres humanos por lo regular pueden plantear preguntas, cuestionar la efectividad o seguridad de las vacunas, o expresar incertidumbre sobre los beneficios y riesgos asociados, lo que se refleja en las emociones negativas presentes (**Expresión de preocupaciones o dudas**).

4.4.6.2 Análisis Psicológico

- Puede reflejar preocupaciones y miedos relacionados con las vacunas. Los individuos pueden experimentar ansiedad, temor o inseguridad en relación con los posibles efectos secundarios, la efectividad de las vacunas o la confianza en los sistemas de salud (***Preocupaciones y miedos***).
- Puede indicar un sentimiento de descontento y desconfianza hacia las vacunas. Los seres humanos en ocasiones pueden expresar frustración, desilusión o escepticismo en relación con la información, la eficacia o la seguridad de las vacunas, lo que se refleja en las emociones negativas expresadas (***Descontento y desconfianza***).
- Puede derivar de experiencias negativas previas con las vacunas. Las personas pueden compartir testimonios de eventos adversos, efectos secundarios o percepciones de daño causado por las vacunas, lo que genera emociones negativas relacionadas (***Experiencias negativas previas***).
- Puede mostrar una discrepancia o crítica hacia las políticas, la comunicación o la implementación de las vacunas. Las personas pueden expresar desacuerdo con las decisiones gubernamentales, la falta de transparencia o la falta de atención a las preocupaciones de salud, lo que se refleja en las emociones negativas presentes (***Expresión de discrepancia o críticas***).

4.4.6.3 Análisis Cognitivo

- Puede reflejar una evaluación cognitiva desfavorable de las vacunas. La gente puede tener creencias y percepciones negativas sobre la eficacia, seguridad o beneficios de las vacunas, lo que se refleja en las emociones negativas expresadas (***Evaluación desfavorable***).
- Puede indicar un sesgo cognitivo hacia lo negativo en la interpretación de la información relacionada con las vacunas. Los seres humanos regularmente pueden prestar más atención a los aspectos negativos, exagerar los posibles riesgos o minimizar los beneficios, lo que influye en las emociones negativas asociadas (***Sesgo cognitivo hacia lo negativo***).
- Puede sugerir la presencia de dudas y preocupaciones cognitivas sobre las vacunas. La gente puede tener preguntas sin respuesta, preocupaciones sobre la efectividad o seguridad, o percepciones de riesgos que generan emociones negativas relacionadas (***Dudas y preocupaciones cognitivas***).
- Puede mostrar una aversión o resistencia cognitiva hacia las vacunas. Los que comentan pueden mostrar una actitud de rechazo, oposición o resistencia hacia las vacunas basada en creencias, valores o experiencias previas, lo que se refleja en las emociones negativas expresadas (***Aversión o resistencia cognitiva***).

4.4.7 Ansiedad

La ansiedad es un estado caracterizado por el miedo, temor e inquietud. Puede manifestarse a través de síntomas como sudoración, inquietud, tensión muscular y palpitaciones. Esta reacción suele ser normal frente a situaciones estresantes, como enfrentar un problema en el trabajo, antes de un examen o al tomar decisiones importantes. Aunque la ansiedad puede ser útil al proporcionar un impulso de energía o mejorar la concentración, para aquellas personas que sufren trastornos de ansiedad, el miedo no es temporal y puede resultar abrumador (MedlinePlus, 2021).

4.4.7.1 Análisis Psicolingüístico

- Puede reflejar la presencia de preocupaciones y temores relacionados con las vacunas. Las personas pueden expresar sentimientos de inquietud, intranquilidad o angustia en relación con los posibles riesgos, efectos secundarios o incertidumbres asociadas a las vacunas (***Preocupaciones y temores***).
- Puede indicar una sensación de incertidumbre y falta de seguridad en relación con las vacunas. Los individuos pueden manifestar dudas, inseguridades o desconocimiento sobre la efectividad, la seguridad o los procesos de desarrollo de las vacunas, lo que genera ansiedad y preocupación (***Incertidumbre y falta de seguridad***).
- Sugiere un impacto emocional negativo asociado con las vacunas. Los que comentan pueden experimentar nerviosismo, inquietud o tensión al hablar o pensar en las vacunas, lo que se refleja en el uso de términos relacionados con la ansiedad (***Impacto emocional negativo***).
- puede indicar la expresión de malestar emocional en relación con las vacunas. Los seres humanos algunas veces pueden compartir sus preocupaciones, experiencias de ansiedad o emociones negativas relacionadas, buscando apoyo, información o validación (***Expresión de malestar emocional***).

4.4.7.2 Análisis Psicológico

- Puede reflejar una preocupación y temor relacionados con las vacunas. La gente puede experimentar niveles elevados de ansiedad debido a la percepción de posibles riesgos, efectos secundarios o incertidumbres asociadas a las vacunas (***Preocupación y temor***).
- Sugiere un impacto emocional negativo asociado con las vacunas. Las personas pueden experimentar sentimientos de nerviosismo, inquietud, tensión o angustia al

hablar o pensar en las vacunas, lo que se refleja en el uso de términos relacionados con la ansiedad (***Impacto emocional negativo***).

- Puede indicar un nivel de estrés y preocupación constante en relación con las vacunas. La gente puede estar constantemente preocupados por los posibles efectos adversos, la efectividad o la seguridad de las vacunas, lo que genera una experiencia de ansiedad persistente (***Estrés y preocupación constante***).
- Puede reflejar una necesidad de control y seguridad en relación con las vacunas. Los individuos pueden expresar un deseo de obtener información adicional, buscar garantías o tomar medidas para mitigar su ansiedad y sentirse más seguros en relación con las vacunas (***Necesidad de control y seguridad***).

4.4.7.3 Análisis Cognitivo

- Puede reflejar un procesamiento cognitivo centrado en información amenazante relacionada con las vacunas. Los seres humanos pueden estar prestando atención y procesando principalmente la información negativa, los posibles riesgos y las preocupaciones asociadas a las vacunas (***Procesamiento de información amenazante***).
- Indica una evaluación cognitiva de los riesgos y las consecuencias percibidas de las vacunas. Los que comentan pueden estar sopesando los posibles efectos secundarios, la efectividad y los beneficios de las vacunas, lo que se refleja en su preocupación y ansiedad relacionada (***Evaluación de riesgos y consecuencias***).
- Puede indicar la presencia de sesgos cognitivos negativos en la forma en que se perciben y procesan las vacunas. Las personas pueden tener una tendencia a dar más peso a la información negativa, exagerar los posibles riesgos y subestimar los beneficios, lo que contribuye a su experiencia de ansiedad (***Sesgos cognitivos negativos***).
- Sugiere una sensación de incertidumbre y falta de control en relación con las vacunas. La gente puede experimentar una falta de certeza sobre la información, la efectividad, la seguridad o los procesos de toma de decisiones en torno a las vacunas, generando ansiedad relacionada con estas percepciones de falta de control (***Incertidumbre y falta de control***).

4.4.8 Enfado

El enfado es una emoción que normalmente expresamos a través de la furia, la irritabilidad, el resentimiento y otros estados similares. Por lo general, el enfado o ira no aparece solo, sino que suele venir acompañado de una serie de efectos físicos, como el aumento del ritmo cardíaco, la presión arterial y los niveles de adrenalina y noradrenalina (DefinicionABC, 2017).

4.4.8.1 Análisis Psicolingüístico

- Refleja una expresión de frustración en relación con las vacunas. Las personas pueden sentirse irritados, molestos o enfadados debido a diferentes razones, como desacuerdos con las políticas de vacunación, experiencias negativas previas o percepciones de desinformación (**Consideración de perspectivas múltiples**).
- Sugiere la presencia de emociones negativas asociadas con las vacunas. Los que comentan pueden experimentar ira, indignación o resentimiento en relación con los temas relacionados con las vacunas, lo cual se refleja en su uso de términos relacionados con el enfado. Expresión de frustración (**Emociones negativas asociadas**).
- Puede indicar una actitud polarizada y una disposición a confrontar en el debate sobre las vacunas. Las personas pueden expresar fuertes opiniones contrarias a las vacunas, mostrando su enfado como una forma de argumentar o defender sus puntos de vista (**Opiniones polarizadas y confrontación**).
- Puede mostrar experiencias de injusticia o descontento percibidas en relación con las vacunas. La gente puede sentir que sus derechos, creencias o valores están siendo infringidos, generando enfado y resentimiento hacia el tema (**Experiencias de injusticia o descontento**).

4.4.8.2 Análisis Psicológico

- Emite una experiencia subjetiva de frustración y disgusto relacionados con las vacunas. Los que comentan pueden sentirse irritados, molestos o enojados debido a diferentes motivos, como desconfianza en las vacunas, desacuerdos con las políticas de vacunación o percepción de problemas relacionados con la seguridad o eficacia de las vacunas (**Frustración y disgusto**).
- Puede indicar un sentimiento de resentimiento y percepción de injusticia en relación con las vacunas. Las personas pueden experimentar un sentido de agravio, percibiendo que sus derechos, creencias o valores están siendo afectados por el tema de las vacunas (**Resentimiento y sentimiento de injusticia**).
- Sugiere una reacción emocional ante una amenaza percibida relacionada con las vacunas. Los que comentan pueden interpretar las vacunas como una amenaza para su salud, libertad o autonomía, generando emociones de enfado como una respuesta a esta percepción de amenaza (**Reacción emocional ante la amenaza percibida**).
- Muestra una expresión de descontento y desacuerdo en relación con las vacunas. Los comentaristas pueden mostrar su malestar y descontento mediante expresiones de enfado, manifestando su oposición o crítica hacia las vacunas y los aspectos asociados (**Expresión de descontento y desacuerdo**).

4.4.8.3 Análisis Cognitivo

- Sugiere una evaluación cognitiva negativa de las vacunas. Los individuos pueden tener creencias negativas arraigadas sobre la seguridad, eficacia o impacto de las vacunas, lo cual influye en su estado emocional de enfado Evaluación negativa de las vacunas (**Consideración de perspectivas múltiples**).
- Puede indicar la presencia de sesgos cognitivos negativos en la forma en que se procesa la información relacionada con las vacunas. Las personas pueden tener una tendencia a centrarse en aspectos negativos, exagerar los riesgos o interpretar la información de manera selectiva, lo que contribuye a su experiencia cognitiva de enfado (**Sesgos cognitivos negativos**).
- Puede reflejar un pensamiento polarizado en relación con las vacunas. Los individuos pueden tener una tendencia a adoptar posturas extremas y rígidas, sin considerar matices o puntos de vista diferentes, lo que se refleja en su expresión de enfado en los comentarios (**Pensamiento polarizado**).
- Puede indicar una resistencia cognitiva al cambio de creencias en relación con las vacunas. Los individuos pueden tener una mentalidad cerrada y una falta de disposición para considerar información nueva o contradictoria que desafíe sus creencias existentes, generando frustración y enfado (**Resistencia al cambio de creencias**).

4.4.9 Cuerpo

El organismo humano es un sistema intrincado y altamente estructurado, compuesto por células que colaboran en conjunto para llevar a cabo funciones específicas indispensables para la supervivencia (Manual MSD, 2022).

4.4.9.1 Análisis Psicolingüístico

- Puede indicar una atención y conciencia particular hacia el cuerpo en relación con las vacunas. Los que comentan pueden expresar preocupaciones sobre posibles efectos secundarios físicos, síntomas o reacciones corporales asociadas con las vacunas (**Conciencia corporal**).
- Sugiere que los individuos pueden estar compartiendo sus experiencias físicas personales en relación con las vacunas. Pueden mencionar sensaciones, cambios o percepciones en su cuerpo después de recibir una vacuna, ya sean positivas o negativas (**Experiencias físicas relacionadas**).
- Refleja la conexión que se establece entre el cuerpo y la salud en el contexto de las vacunas. Los seres humanos a veces pueden expresar preocupaciones sobre el

impacto de las vacunas en su bienestar físico y cómo las perciben en relación con su salud general (***Vinculación entre cuerpo y salud***).

- Puede manifestar preocupaciones sobre la seguridad física relacionada con las vacunas. La gente puede expresar temores o dudas sobre los posibles efectos secundarios o riesgos para su cuerpo, lo cual se refleja en su elección de palabras relacionadas con el cuerpo (***Preocupaciones sobre la seguridad física***).

4.4.9.2 Análisis Psicológico

- Sugiere una preocupación por la salud física en relación con las vacunas. Los que comentan pueden expresar temores, inquietudes o interés en cómo las vacunas pueden afectar su cuerpo y su bienestar físico (***Preocupación por la salud física***).
- Puede indicar una mayor sensibilidad a las sensaciones corporales y la capacidad de notar y describir cambios físicos relacionados con las vacunas. Los individuos pueden estar atentos a cualquier síntoma o reacción en su cuerpo después de recibir una vacuna (***Sensibilidad a las sensaciones corporales***).
- Muestra la importancia de las experiencias personales en relación con las vacunas. La gente puede compartir sus propias experiencias físicas, síntomas o reacciones corporales, lo cual refleja su enfoque en cómo las vacunas afectan su propio cuerpo (***Experiencias personales***).
- Indica una conexión entre el cuerpo y el bienestar general. Las personas pueden considerar que su cuerpo es un aspecto importante de su salud y bienestar, y expresar preocupaciones o percepciones en relación con las vacunas que pueden afectar su cuerpo de alguna manera (***Conexión entre cuerpo y bienestar***).

4.4.9.3 Análisis Cognitivo

- Refleja una mayor conciencia y atención hacia las sensaciones y experiencias corporales relacionadas con las vacunas. Los individuos pueden estar más conscientes de los efectos físicos y las sensaciones que perciben en su cuerpo después de recibir una vacuna (***Conciencia corporal***).
- Sugiere que los seres humanos en ocasiones están evaluando cognitivamente los riesgos y beneficios asociados con las vacunas en relación con su cuerpo. Pueden estar sopesando los posibles efectos secundarios y los beneficios esperados en términos de su salud física (***Evaluación cognitiva de riesgos y beneficios***).
- Indica que la gente está procesando y evaluando información específica relacionada con el cuerpo en el contexto de las vacunas. Pueden estar considerando cómo las vacunas afectan directamente su cuerpo y qué implicaciones tiene esto para su salud y bienestar (***Procesamiento de información relacionada con el cuerpo***).
- Sugiere que las personas pueden tener una atención selectiva hacia los aspectos físicos de las vacunas. Pueden estar enfocados en cómo las vacunas interactúan con

su cuerpo, los posibles efectos secundarios físicos o cualquier cambio físico que perciban (*Atención selectiva a aspectos físicos*).

4.4.10 Salud

La salud constituye un pilar esencial para el desarrollo y la existencia humana. En términos simples, puede ser concebida como un estado de bienestar integral que abarca lo físico, lo mental y lo social, trascendiendo la mera ausencia de enfermedad o dolencias físicas. Esta visión amplia de la salud fue introducida por la Organización Mundial de la Salud (OMS) en su constitución de 1948, lo cual refleja la comprensión de la salud como un recurso vital y no meramente como un objetivo de la vida (Clínica Universidad de Navarra, 2022).

4.4.10.1 Análisis Psicolingüístico

- Sugiere una preocupación en relación con el bienestar físico y mental en el contexto de las vacunas. Los que comentan pueden expresar inquietudes, opiniones o información relacionada con cómo las vacunas afectan la salud individual y colectiva (*Preocupación por la salud*).
- Refleja un enfoque en los aspectos médicos y la evaluación de los efectos de las vacunas en la salud. Los individuos pueden mencionar síntomas, efectos secundarios, beneficios o riesgos en relación con la salud al discutir las vacunas (*Enfoque en los aspectos médicos*).
- Puede indicar una búsqueda de información y orientación en relación con las vacunas y sus implicaciones para la salud. Las personas pueden expresar la necesidad de información precisa y confiable sobre cómo las vacunas pueden impactar la salud (*Búsqueda de información y orientación*).
- Puede reflejar la expresión de preferencias y valores en relación con la toma de decisiones sobre las vacunas. La gente puede considerar la salud como un factor clave al evaluar los beneficios y riesgos de las vacunas y al expresar sus posturas (*Expresión de preferencias y valores*).

4.4.10.2 Análisis Psicológico

- Refleja la importancia que las personas asignan a la salud en el contexto de las vacunas. Puede indicar una preocupación por el bienestar físico y mental, y la percepción de que las vacunas tienen un impacto directo en la salud individual y colectiva (*Importancia de la salud*).

- Sugiere que los individuos están evaluando cognitivamente los posibles riesgos y beneficios de las vacunas en relación con la salud. Pueden considerar los efectos positivos en la prevención de enfermedades, así como los posibles efectos secundarios o riesgos para la salud (***Evaluación de riesgos y beneficios***).
- Puede indicar la confianza en las intervenciones de salud, como las vacunas, para mantener y mejorar la salud. Los que comentan pueden expresar la creencia en la eficacia de las vacunas y su contribución a la prevención de enfermedades y promoción de la salud (***Confianza en las intervenciones de salud***).
- Puede reflejar emociones y preocupaciones relacionadas con la salud en el contexto de las vacunas. Las personas pueden expresar miedo a enfermedades, ansiedad por los posibles efectos secundarios o frustración por la percepción de riesgos para la salud (***Emociones y preocupaciones relacionadas con la salud***).

4.4.10.3 Análisis Cognitivo

- Sugiere que los individuos están prestando atención selectiva a los aspectos relacionados con la salud en el contexto de las vacunas. Pueden estar enfocados en evaluar y procesar información específica sobre cómo las vacunas impactan la salud individual y colectiva (***Atención selectiva***).
- Indica que los seres humanos con normal frecuencia están evaluando cognitivamente los posibles riesgos y beneficios de las vacunas en relación con la salud. Pueden estar sopesando los efectos positivos en la prevención de enfermedades y los posibles efectos secundarios o riesgos para la salud al tomar decisiones sobre las vacunas (***Evaluación de riesgos y beneficios***).
- Refleja una conexión cognitiva entre las vacunas y el bienestar general. La gente puede percibir las vacunas como una medida importante para mantener y promover la salud, y están considerando cómo las vacunas pueden afectar su bienestar físico y mental (***Conexión entre vacunas y bienestar***).
- Sugiere que la gente posee información y conocimiento sobre aspectos relacionados con la salud y las vacunas. Pueden expresar opiniones fundamentadas y basadas en su comprensión cognitiva de cómo las vacunas afectan la salud (***Información y conocimiento sobre la salud***).

4.4.11 Muerte

La muerte representa el término de la vida o su interrupción, dependiendo de cómo se enfoque. Es un evento inevitable que compartimos los seres humanos con todas las formas de vida, si bien cada una tiene su propio ciclo de existencia. Constituye la experiencia final para todos los organismos vivos, aunque puede ocurrir en momentos distintos. En ocasiones, se debe a circunstancias vitales, como encuentros

con depredadores o participación en desastres naturales, mientras que en otros casos se debe simplemente a enfermedades y el deterioro del organismo (Concepto, 2023).

4.4.11.1 Análisis Psicolingüístico

- Sugiere que los comentaristas pueden tener temor o preocupación acerca de los posibles riesgos asociados con las vacunas. Puede indicar una percepción de que las vacunas podrían tener efectos adversos graves que podrían poner en peligro la vida. **(Temor o preocupación).**
- Puede reflejar la expresión de riesgo percibido en relación con las vacunas. Los individuos pueden estar discutiendo o expresando su opinión sobre los posibles efectos negativos en términos de mortalidad asociada a las vacunas **(Expresión de riesgo percibido).**
- Puede indicar un enfoque en la seguridad y los posibles efectos secundarios graves relacionados con las vacunas. Los que comentan pueden estar debatiendo sobre la relación entre las vacunas y la mortalidad, y pueden mencionar casos o situaciones específicas donde se haya reportado la muerte como resultado de la vacunación **(Debate sobre seguridad y efectos secundarios).**
- Puede reflejar una expresión emocional intensa en relación con las vacunas. Los individuos pueden estar transmitiendo preocupación, miedo o indignación ante la posibilidad de que las vacunas puedan estar asociadas con la muerte **(Expresión emocional intensa).**

4.4.11.2 Análisis Psicológico

- Sugiere que las personas pueden experimentar temor o ansiedad en relación con las vacunas. Puede reflejar una preocupación por los posibles efectos negativos en términos de mortalidad asociada a las vacunas **(Temor o ansiedad).**
- Indica una preocupación por la seguridad en relación con las vacunas. La gente puede expresar inquietudes sobre posibles riesgos y efectos adversos graves, incluyendo la posibilidad de muerte, al discutir las vacunas **(Preocupación por la seguridad).**
- Puede indicar que los individuos están emocionalmente afectados por el tema de las vacunas y la posibilidad de consecuencias fatales. Puede reflejar una respuesta emocional intensa, como miedo, preocupación o indignación **(Impacto emocional).**
- Sugiere que los que comentan están evaluando cognitivamente los posibles riesgos y beneficios de las vacunas, incluyendo el riesgo de muerte. Pueden estar sopesando los posibles efectos negativos en términos de mortalidad en comparación con los beneficios de la vacunación **(Evaluación de riesgos y beneficios).**

4.4.11.3 Análisis Cognitivo

- Sugiere que las personas están evaluando cognitivamente los posibles riesgos asociados con las vacunas. Pueden estar considerando la probabilidad de consecuencias graves, como la muerte, al tomar decisiones sobre la vacunación (**Evaluación de riesgos**).
- Indica una focalización cognitiva en resultados extremos y potencialmente negativos. Los individuos pueden estar preocupados por los efectos adversos más graves y están expresando estas preocupaciones en sus comentarios (**Focalización en resultados extremos**).
- Sugiere que las personas consideran la posibilidad de muerte como un factor relevante y significativo en sus discusiones sobre las vacunas. Pueden estar ponderando los riesgos y beneficios de las vacunas y considerando el impacto potencial en términos de pérdida de vida (**Valoración de la importancia**).
- Refleja una sensibilidad cognitiva hacia la seguridad en el contexto de las vacunas. Los seres humanos pueden estar prestando atención especial a los posibles riesgos y peligros asociados con las vacunas, incluyendo la posibilidad de muerte, como parte de su evaluación general (**Sensibilidad a la seguridad**).

4.4.12 Signos de Puntuación, Interrogación y Exclamación

4.4.12.1 Análisis Psicolingüístico

- Sugieren una expresión intensa de emociones en los comentarios. Pueden reflejar sorpresa, indignación, entusiasmo, enfado u otras emociones fuertes relacionadas con el tema de las vacunas (**Intensidad emocional**).
- El uso abundante de signos de interrogación indica que los que comentan plantean muchas preguntas y dudas en relación con las vacunas. Puede reflejar una búsqueda de información, inseguridad o inquietud sobre diversos aspectos relacionados con la vacunación (**Expresión de preguntas y dudas**).
- El uso de signos de puntuación, como la exclamación, puede indicar una fuerte enfatización de ciertas ideas o afirmaciones en los comentarios. Los individuos pueden querer destacar ciertos puntos y transmitir una mayor convicción o convencimiento (**Enfatización de ideas**).
- El uso de signos de interrogación y exclamación puede indicar sorpresa o asombro por determinadas afirmaciones o información relacionada con las vacunas. Puede reflejar reacciones emocionales inmediatas a los argumentos o datos presentados en los comentarios (**Expresión de sorpresa o asombro**).

4.4.12.2 Análisis Psicológico

- El uso frecuente de signos de puntuación, especialmente de interrogación y exclamación, indica una expresión emocional intensa en los comentarios. Puede reflejar emociones como sorpresa, entusiasmo, indignación o enfado relacionadas con el tema de las vacunas. Los individuos pueden estar expresando fuertes sentimientos y reacciones emocionales ante las opiniones o información compartida (***Intensidad emocional***).
- El uso de signos de puntuación puede indicar una implicación personal por parte de las personas en el tema de las vacunas. Pueden estar compartiendo sus propias experiencias, opiniones o creencias de manera enérgica y apasionada. Los signos de puntuación intensifican la expresión y dan un mayor énfasis a su punto de vista (***Implicación personal***).
- El uso de signos de interrogación puede indicar incredulidad o escepticismo hacia ciertas afirmaciones o información relacionada con las vacunas. Los seres humanos en algunas ocasiones pueden estar expresando dudas o cuestionando la veracidad de ciertos argumentos o datos. Los signos de puntuación enfatizan su incredulidad y refuerzan su expresión de escepticismo (***Incredulidad o escepticismo***).
- El uso de signos de exclamación puede indicar sorpresa, asombro o admiración ante ciertos aspectos relacionados con las vacunas. La gente puede estar reaccionando de manera emotiva a información positiva o negativa y expresando su asombro o admiración en los comentarios. Los signos de puntuación refuerzan esta expresión emocional (***Expresión de sorpresa o admiración***).

4.4.12.3 Análisis Cognitivo

- El uso frecuente de signos de puntuación puede indicar un mayor nivel de atención y enfoque en los comentarios. Los que comentan pueden estar expresando ideas de manera más clara y organizada al utilizar signos de puntuación para delimitar frases y separar ideas. Esto puede reflejar una mayor capacidad cognitiva para estructurar y comunicar sus pensamientos (***Atención y enfoque***).
- El uso de signos de puntuación, especialmente de interrogación y exclamación, pueden indicar una intensificación en la forma en que se expresan las ideas. Los individuos pueden estar resaltando ciertos puntos, enfatizando sus argumentos o mostrando un mayor grado de convicción. Esto sugiere una mayor implicación cognitiva y una intención de transmitir de manera más enfática sus opiniones (***Intensificación de la expresión***).
- El uso de signos de interrogación indica que los comentaristas están planteando preguntas y buscando respuestas en relación con las vacunas. Esto refleja una actividad cognitiva orientada hacia la adquisición de conocimiento y la búsqueda de información relevante. Los signos de interrogación pueden indicar una actitud

cognitiva abierta y un deseo de explorar diferentes perspectivas (**Curiosidad y búsqueda de respuestas**).

- El uso de signos de exclamación puede indicar una mayor intensidad emocional en los comentarios. Esto puede estar relacionado con la activación cognitiva de emociones intensas, como sorpresa, asombro o indignación, en respuesta a ciertos aspectos de las vacunas. Los signos de exclamación enfatizan la expresión emocional y refuerzan la carga cognitiva asociada a estas emociones (**Reacciones emocionales intensas**).

Haciendo un breve resumen se puede decir que fueron aplicados los modelos J48, Bayes Net y SMO para analizar las características psicolingüísticas más relevantes y al utilizar el método de ganancia de información, se identificaron las siguientes características con mayor peso: "pronombres personales", "negación", "verbos", "familia", "emociones positivas", "emociones negativas", "ansiedad", "enfado", "cuerpo", "salud", "muerte" y signos de "interrogación", "exclamación" y "puntuación".

	Análisis		
	Psicolingüístico	Psicológico	Cognitivo
Pronombres Personales	<ul style="list-style-type: none"> **Autoenfoque **Personalización del discurso **Identificación y pertenencia 	<ul style="list-style-type: none"> **Autoafirmación y expresión de identidad **Apego a creencias y valores **Empatía y conexión emocional **Enfoque en la responsabilidad individual 	<ul style="list-style-type: none"> **Enfoque en la experiencia personal **Identificación subjetiva **Autoafirmación y automeferencia
Negación	<ul style="list-style-type: none"> **Expresión de dudas o escepticismo **Contrarrestar argumentos contrarios **Manifestación de desacuerdo o disconformidad 	<ul style="list-style-type: none"> **Resistencia al cambio **Mecanismo de defensa **Necesidad de coherencia cognitiva 	<ul style="list-style-type: none"> **Diferencias en la valoración de la información **Evaluación de riesgos y beneficios **Defensa de creencias previas
verbos	<ul style="list-style-type: none"> **Expresión de acciones y opiniones **Manifestación de compromiso y participación **Enfoque en la acción y el cambio 	<ul style="list-style-type: none"> **Orientación hacia la acción **Manifestación de actitudes y emociones **Expresión de control y empoderamiento 	<ul style="list-style-type: none"> **Procesamiento activo de la información **Enfoque en la acción y los resultados **Evaluación de causas y consecuencias
Familia	<ul style="list-style-type: none"> **Valoración de la influencia familiar **Experiencias compartidas **Perspectiva intergeneracional **Apoyo y resistencia familiar 	<ul style="list-style-type: none"> **Influencia social y emocional **Identificación y pertenencia **Responsabilidad y protección **Conflicto y diferencias de opinión 	<ul style="list-style-type: none"> **Asociaciones cognitivas y conexiones personales **Consideración de perspectivas múltiples **Evaluación de beneficios y riesgos **Construcción de identidad y roles familiares
Emociones Positivas	<ul style="list-style-type: none"> **Actitud favorable **Validación emocional **Compartir experiencias positivas **Construcción de confianza y optimismo 	<ul style="list-style-type: none"> **Actitud favorable y satisfacción **Validación y apoyo emocional **Consideración de perspectivas múltiples **Reforzamiento de creencias y actitudes 	<ul style="list-style-type: none"> **Evaluación favorable de las vacunas **Atención selectiva a aspectos positivos **Creencias optimistas y esperanzadoras **Sesgo cognitivo hacia lo positivo
Emociones Negativas	<ul style="list-style-type: none"> **Actitudes desfavorables **Expresión de frustración o descontento **Validación emocional de experiencias negativas **Expresión de preocupaciones o dudas 	<ul style="list-style-type: none"> **Preocupaciones y miedos **Descontento y desconfianza **Experiencias negativas previas **Expresión de discrepancia o críticas 	<ul style="list-style-type: none"> **Evaluación desfavorable **Sesgo cognitivo hacia lo negativo **Dudas y preocupaciones cognitivas **Aversión o resistencia cognitiva
Ansiedad	<ul style="list-style-type: none"> **Preocupaciones y temores **Incertidumbre y falta de seguridad **Impacto emocional negativo **Expresión de malestar emocional 	<ul style="list-style-type: none"> **Preocupación y temor **Estrés y preocupación constante **Necesidad de control y seguridad 	<ul style="list-style-type: none"> **Procesamiento de información amenazante **Evaluación de riesgos y consecuencias **Sesgos cognitivos negativos **Incertidumbre y falta de control
Enfado	<ul style="list-style-type: none"> **Consideración de perspectivas múltiples **Emociones negativas asociadas **Opiniones polarizadas y confrontación **Experiencias de injusticia o descontento 	<ul style="list-style-type: none"> **Frustración y disgusto **Resentimiento y sentimiento de injusticia **Reacción emocional ante la amenaza percibida **Expresión de descontento y desacuerdo 	<ul style="list-style-type: none"> **Consideración de perspectivas múltiples **Sesgos cognitivos negativos **Pensamiento polarizado **Resistencia al cambio de creencias
Cuerpo	<ul style="list-style-type: none"> **Conciencia corporal **Experiencias físicas relacionadas **Vinculación entre cuerpo y salud **Preocupaciones sobre la seguridad física 	<ul style="list-style-type: none"> **Preocupación por la salud física **Sensibilidad a las sensaciones corporales **Experiencias personales **Conexión entre cuerpo y bienestar 	<ul style="list-style-type: none"> **Evaluación cognitiva de riesgos y beneficios **Procesamiento de información relacionada con el cuerpo **Atención selectiva a aspectos físicos
Salud	<ul style="list-style-type: none"> **Preocupación por la salud **Enfoque en los aspectos médicos **Búsqueda de información y orientación **Expresión de preferencias y valores 	<ul style="list-style-type: none"> **Importancia de la salud **Evaluación de riesgos y beneficios **Confianza en las intervenciones de salud **Emociones y preocupaciones relacionadas con la salud 	<ul style="list-style-type: none"> **Atención selectiva **Evaluación de riesgos y beneficios **Conexión entre vacunas y bienestar **Información y conocimiento sobre la salud
Muerte	<ul style="list-style-type: none"> **Temor o preocupación **Expresión de riesgo percibido **Debate sobre seguridad y efectos secundarios **Expresión emocional intensa 	<ul style="list-style-type: none"> **Temor o ansiedad **Preocupación por la seguridad **Impacto emocional **Evaluación de riesgos y beneficios 	<ul style="list-style-type: none"> **Evaluación de riesgos **Focalización en resultados extremos **Valoración de la importancia **Sensibilidad a la seguridad
Period	<ul style="list-style-type: none"> **Intensidad emocional 	<ul style="list-style-type: none"> **Implicación personal 	<ul style="list-style-type: none"> **Atención y enfoque
QMark	<ul style="list-style-type: none"> **Expresión de preguntas y dudas 	<ul style="list-style-type: none"> **Incredulidad o escepticismo 	<ul style="list-style-type: none"> **Intensificación de la expresión
Exclam	<ul style="list-style-type: none"> **Enfatización de ideas **Expresión de sorpresa o asombro 	<ul style="list-style-type: none"> **Expresión de sorpresa o admiración 	<ul style="list-style-type: none"> **Curiosidad y búsqueda de respuestas **Reacciones emocionales intensas

Estas características representan aspectos importantes en el análisis de los datos analizados (corpus) con las posturas provacunas y antivacunas. Los pronombres personales pueden revelar la subjetividad de los mensajes, mientras que la negación puede indicar la oposición a ciertas ideas. Los verbos son fundamentales para expresar acciones y estados, y la familia puede estar relacionada con la percepción de responsabilidad y cuidado.

Las emociones positivas y negativas son relevantes en la comprensión de las actitudes hacia las vacunas, al igual que la ansiedad y el enfado, que reflejan estados emocionales intensos. El cuerpo y la salud son conceptos centrales en el contexto de la vacunación, y la muerte representa el temor a las consecuencias negativas. Los signos de interrogación, exclamación y puntuación pueden transmitir énfasis y emociones adicionales en los mensajes.

Por lo tanto, estas características psicolingüísticas destacadas proporcionan información valiosa sobre los discursos provacunas y antivacunas, ayudando a comprender mejor las actitudes, emociones y creencias asociadas con la vacunación.

CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES

Las investigaciones sobre los factores que determinan la fuerza de los movimientos antivacunas en las redes sociales son esenciales para desarrollar estrategias efectivas de contrarresto, informar políticas de salud pública y comprender el impacto de estos movimientos en la sociedad. Sin embargo, las redes sociales pueden proporcionar una plataforma efectiva para que los movimientos antivacunas difundan su mensaje y alcancen a un público amplio.

La viralidad de la información en las redes sociales puede dificultar la tarea de contrarrestar la desinformación antivacunas. La velocidad y facilidad con la que se comparten contenidos en línea pueden hacer que los mensajes falsos o engañosos se difundan más rápidamente que la información precisa y basada en evidencia.

Es fundamental que los profesionales de la salud, las autoridades sanitarias y las plataformas de redes sociales trabajen en conjunto para abordar la desinformación y promover una comunicación clara y basada en evidencia sobre las vacunas en estos espacios digitales.

Las investigaciones sobre los factores que determinan la fuerza de los movimientos antivacunas en las redes sociales son esenciales para desarrollar estrategias efectivas de contrarresto, informar políticas de salud pública y comprender el impacto de estos movimientos en la sociedad. Estas investigaciones proporcionan una base sólida para abordar los desafíos planteados por los movimientos antivacunas y promover la vacunación.

En esta tesis de acuerdo al análisis de características desde un punto de vista Psicolingüístico, psicológico y cognitivo se pueden obtener los siguientes factores de rechazo que determina a los movimientos antivacunas:

- **Apego a creencias y valores:** Las personas pueden aferrarse a creencias y valores que cuestionan o rechazan la vacunación, basándose en principios ideológicos o religiosos.
- **Resistencia al cambio:** Algunas personas pueden resistirse a los cambios en las prácticas de salud pública, como la vacunación obligatoria, y pueden manifestar oposición como resultado.
- **Necesidad de coherencia cognitiva:** Las personas pueden tener una necesidad de mantener una coherencia entre sus creencias y actitudes, lo que puede llevar a la negación o el rechazo de la información que contradice sus puntos de vista antivacunas.
- **Influencia social y emocional:** La influencia de grupos o comunidades que se oponen a las vacunas puede desempeñar un papel importante en la adhesión a los movimientos antivacunas. La conexión emocional y la identificación con estos grupos pueden reforzar la resistencia a la vacunación.
- **Preocupaciones y miedos:** Las preocupaciones relacionadas con los posibles efectos secundarios, la seguridad de las vacunas y los temores infundados pueden contribuir a la formación de movimientos antivacunas.
- **Experiencias personales negativas:** Las experiencias negativas previas con la vacunación, como reacciones adversas raras, pueden generar desconfianza y contribuir a la resistencia a las vacunas.
- **Descontento y desconfianza:** Las actitudes negativas hacia las autoridades sanitarias y la industria farmacéutica, así como la desconfianza en los sistemas de atención médica, pueden impulsar la adhesión a los movimientos antivacunas.
- **Expresión de control y empoderamiento:** Algunas personas pueden encontrar una sensación de control y empoderamiento al resistirse a las vacunas y tomar decisiones de salud independientes.
- **Evaluación de riesgos y beneficios:** La percepción de riesgos y beneficios puede estar sesgada, y algunas personas pueden sobrestimar los riesgos asociados con las vacunas mientras subestiman los beneficios.

Es importante tener en cuenta que no todas las personas que expresan preocupaciones o dudas sobre las vacunas en las redes sociales son necesariamente parte de un movimiento antivacunas organizado. Sin embargo, los movimientos antivacunas en línea representan un componente significativo del panorama de la información sobre vacunas en las redes sociales y pueden influir en las actitudes y decisiones individuales.

CAPÍTULO VI
PRODUCTOS ACADÉMICOS

REFERENCIAS

- Artmann, P. (2020). *¿Qué es un verbo?* Arbolabc.Com. <https://arbolabc.com/recursos/verbos#:~:text=Los verbos son aquellas palabras,completa sin usar un verbo.>
- AutoGenesis. (2023). *LAS EMOCIONES POSITIVAS.* <https://www.autognosis.com/psicologia/emociones-positivas/#:~:text=Las emociones positivas son aquellas,un periodo corto de tiempo.>
- Clinica Universidad de Navarra. (2022). *Salud.* <https://www.cun.es/diccionario-medico/terminos/salud#:~:text=En su definición más sencilla,de enfermedad o afecciones físicas.>
- Concepto. (2023). *¿Qué es la muerte?* <https://concepto.de/muerte/>
- decide. (2019). *Procesamiento del Lenguaje Natural (PLN o NLP): qué es y para qué se utiliza.* <https://decidesoluciones.es/procesamiento-del-lenguaje-natural-pln-o-nlp-que-es-y-para-que-se-utiliza/>
- DefinicionABC. (2017). *Definición de Enfado.* <https://www.definicionabc.com/general/enfado.php>
- Fernández Rivera, L. (2019). *La diferencia entre Social Media y Redes Sociales.* Mayo 2. <https://www.planning.com.ar/la-diferencia-entre-social-media-y-redes-sociales/>
- GRAPH Everywhere. (2022). *Machine Learning | Qué es, tipos, ejemplos y cómo implementarlo.* <https://www.grapheverywhere.com/machine-learning-que-es-tipos-ejemplos-y-como-implementarlo/>
- IBERDROLA. (2022). *QUÉ ES EL “MACHINE LEARNING.”* <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico>
- IBM. (2021). *Conceptos básicos de ayuda de CRISP-DM.* Copyright IBM Corporation 1994, 2021. <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- IEPP - Instituto Europeo de EmocionesPositivas. (2023). *Emociones Negativas.* <https://www.iepp.es/emociones-negativas/#:~:text=Una emocion negativa es aquella,y en su propia vida.>
- Learn Microsoft. (2022). *Conceptos de minería de datos.* <https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>
- López Santamaría, M. Á. (2015). *Los movimientos antivacunacion y su presencia*

en internet. https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1988-348X2015000300011

Manual MSD. (2022). *Introducción al cuerpo humano*.

MedlinePlus. (2021). *Ansiedad*.
<https://medlineplus.gov/spanish/anxiety.html#:~:text=¿Qué es la ansiedad%3F,una reacción normal al estrés.>

NextVisiion. (2021). *¿Que tipos de Redes Sociales existen?*
<https://nextvision.mx/blog/social-media/tipos-de-redes-sociales>

Pérez Porto, J., Gardey, A. (2023). *DEFINICIÓN DE NEGACIÓN*. Definicion.DE.
<https://definicion.de/negacion/>

Sergio Galán. (2016). *Medios Sociales ¿Qué son?* <https://sergiogalang.com/que-son-medios-sociales/>

Significados. (2019). *Familia*. Significados.Com.
<https://www.significados.com/familia/>

UPB. (2020). *UPB Virtual. ¿Qué Significan y Cuáles Son Los Pronombres Personales?* <https://www.upb.edu.co/es/central-blogs/ortografia/pronombres-personales#:~:text=Los pronombres personales son palabras,es quien emite el mensaje.>

VIEWNEXT. (2020). *¿Qué es el PLN (Procesamiento del Lenguaje Natural)*. 05 Marzo. <https://www.viewnext.com/que-es-pln-procesamiento-lenguaje-natural/>