



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería de Sistemas

**Plataforma web para el análisis de sentimiento en
comentarios de YouTube como herramienta de
evaluación de mercado en PYMEs utilizando Machine
Learning**

TESIS

Para optar el Título Profesional de Ingeniero de Sistemas

AUTOR

Dina Andrea DE LA JARA CORDERO

ASESOR

Dra. Rosa Sumactika DELGADILLO AVILA

Lima, Perú

2025



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

De la Jara, D. (2025). *Plataforma web para el análisis de sentimiento en comentarios de YouTube como herramienta de evaluación de mercado en PYMEs utilizando Machine Learning*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Sistemas]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios autor/ asesor

Datos de autor	
Nombres y apellidos	Dina Andrea de la Jara Cordero
Tipo de documento de identidad	DNI
Número de documento de identidad	71833752
URL de ORCID	https://orcid.org/0009-0008-1064-0702
Datos de asesor	
Nombres y apellidos	Rosa Sumactika Delgadillo Ávila
Tipo de documento de identidad	DNI
Número de documento de identidad	06445553
URL de ORCID	https://orcid.org/0000-0003-4899-3008
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Jorge Leoncio Guerra Guerra
Tipo de documento	DNI
Número de documento de identidad	08473333
Miembro del jurado 1	
Nombres y apellidos	Félix Armando Fermín Pérez
Tipo de documento	DNI
Número de documento de identidad	08736347
Datos de investigación	
Línea de investigación	Inteligencia Artificial

Grupo de investigación	Inteligencia Artificial
Agencia de financiamiento	No aplica
Ubicación geográfica de la investigación	Lima, Perú
Año o rango de años en que se realizó la investigación	2024–2025
URL de disciplinas OCDE	2.02.04 -- Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04



**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Profesional de Ingeniería de Sistemas**

Acta Virtual de Sustentación de Tesis

Siendo las 10:00 horas del día 08 de agosto del año 2025, se reunieron virtualmente los docentes designados como miembros de Jurado de Tesis, presidido por el Mg. Jorge Guerra Guerra, Ing. Felix Armando Fermin Pérez (Miembro) y la Dra. Rosa Delgadillo Avila (Miembro Asesor), usando la plataforma Meet (<https://meet.google.com/rip-ogaq-vmk>), para la sustentación Virtual de la tesis Intitulada: **"PLATAFORMA WEB PARA EL ANÁLISIS DE SENTIMIENTO EN COMENTARIOS DE YOUTUBE COMO HERRAMIENTA DE EVALUACIÓN DE MERCADO EN PYMES UTILIZANDO MACHINE LEARNING"**, de la Bachiller: **Dina Andrea DE LA JARA CORDERO**; para obtener el Título Profesional de Ingeniero de Sistemas.

Acto seguido de la exposición de la Tesis, el Presidente invitó al Bachiller a responder las preguntas formuladas por los Miembros del Jurado.

El Bachiller, en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez las preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, el bachiller obtuvo la nota de dieciseis **16**

A continuación, el Presidente del Jurado Mg. Jorge Guerra Guerra, declara al Bachiller **Ingeniero de Sistemas**.

Siendo 11.09 horas, se levantó la sesión.

Mg. Jorge Guerra Guerra
Presidente

Ing. Felix Armando Fermin Pérez
Miembro

Dra. Rosa Delgadillo Avila
Miembro Asesor

Anexo 1

CERTIFICADO DE SIMILITUD

Yo, **Rosa Sumactika DELGADILLO AVILA** en mi condición de asesor acreditado con el Dictamen de la EP N°000009-2024-R-EPISI-FISI/UNMSM de la **tesis, c u y o** título es **Plataforma Web para el Análisis de Sentimiento en comentarios de YouTube como herramienta de evaluación de mercado en PYMEs utilizando Machine Learning**, presentado por el **bachiller Dina Andrea DE LA JARA CORDERO** para optar el título de **INGENIERO DE SISTEMAS**, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el 14% de similitud, nivel PERMITIDO para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional**.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del título profesional de Ingeniero de Sistemas.

Lima, 16 de setiembre 2025

Firma del Asesor



Nombres y Apellidos del asesor: Rosa Sumactika Delgadillo Avila de Mauricio

DNI:06445553

* Según corresponda: bachiller, magíster, doctor; título profesional o título de segunda especialidad

La vigencia del presente certificado corresponde a un período de noventa (90) días desde la fecha de su emisión; siempre que no haya observaciones en el texto durante el proceso de revisión por el jurado informante o surja alguna contingencia.

DEDICATORIA

Este trabajo de investigación está dedicado a mis padres, Lourdes y Carlos, quienes son el faro que ilumina mi camino, mi fuerza constante y el motivo por el que me esfuerzo cada día. Gracias por su amor incondicional, su sabiduría y por ser mi mayor inspiración en cada etapa de mi vida.

AGRADECIMIENTOS

En primer lugar, agradezco a Dios por acompañarme en cada paso, ser mi guía y por darme la fortaleza necesaria en los momentos más difíciles.

A mis padres y hermanos, por su fe inquebrantable en mí, su apoyo constante y por enseñarme que, al estar juntos, no hay sueño que no podamos alcanzar.

A la profesora Dra. Rosa Delgadillo Ávila, por su valiosa orientación, su paciencia y por cada consejo que marcó mi camino, no solo en el desarrollo de esta tesis, sino también a lo largo de toda mi carrera.

A la Universidad Nacional Mayor de San Marcos y a la Facultad de Ingeniería de Sistemas, por brindarme una formación sólida y valores que hoy me permiten culminar esta etapa con gratitud.

RESUMEN

En la actualidad, la creciente digitalización ha generado un notable incremento en el uso de redes sociales como una fuente de intercambio de información y opinión. En este contexto, el análisis de sentimientos se ha convertido en una herramienta clave para la toma de decisiones empresariales al permitir evaluar la percepción de los clientes sobre un producto o servicio a partir de sus comentarios en redes sociales.

Sin embargo, muchas pequeñas y medianas empresas (PYMEs) en el Perú no cuentan con acceso a este tipo de tecnologías, debido a factores como el desconocimiento del potencial del análisis de datos, falta de recursos económicos, ausencia de personal especializado o limitaciones en su infraestructura tecnológica. Ante esta problemática, la presente investigación propone el diseño y desarrollo de una plataforma web para el análisis de sentimientos de comentarios en español de videos de YouTube como herramienta de apoyo en la evaluación de la percepción del mercado para las PYMEs.

La plataforma web propuesta implementa modelos de Machine Learning y Procesamiento de Lenguaje Natural (NLP), tales como VADER, TextBlob y BETO para clasificar los comentarios en español según su sentimiento: positivos, negativos o neutros. El sistema se construyó bajo la arquitectura Modelo-Vista-Controlador (MVC) e integra la metodología CRISP-DM para la minería de datos junto con la metodología Scrum para el desarrollo ágil de la plataforma web.

Los resultados experimentales, obtenidos a partir del análisis de 3 000 comentarios de diferentes videos, evidenciaron que, tras comparar los tres modelos evaluados, BETO obtuvo una exactitud del 81,30% consolidándose como el modelo principal para el análisis de sentimientos en la investigación.

Palabras clave: análisis de sentimientos, percepción del cliente, evaluación de mercado, machine learning, YouTube, PYMEs, procesamiento de lenguaje natural, BETO, VADER, TextBlob, CRISP-DM, Scrum, Python.

ABSTRACT

Currently, increasing digitalization has generated a remarkable rise in the use of social networks as a means of sharing information and opinions. In this context, sentiment analysis has become a key tool for business decision-making, as it allows evaluating customer perceptions of a product or service based on their comments on social networks.

However, many small and medium-sized enterprises (SMEs) in Peru do not have access to this type of technology due to factors such as lack of awareness of the potential of data analysis, limited economic resources, absence of specialized personnel, or constraints in technological infrastructure. To address this issue, this research proposes the design and development of a web platform for sentiment analysis of Spanish-language comments on YouTube videos as a support tool for assessing market perception for SMEs.

The proposed web platform implements Machine Learning and Natural Language Processing (NLP) models such as VADER, TextBlob, and BETO, classifying comments in Spanish into positive, negative, or neutral categories. The system was developed using the Model-View-Controller (MVC) architecture and integrates the CRISP-DM methodology for data mining along with the Scrum methodology for the agile development of the web platform.

The experimental results, obtained from the analysis of 3 000 comments from different videos, showed that, after comparing the performance of the three evaluated models, BETO achieved an accuracy of 81.30%, establishing itself as the main model proposed for sentiment analysis in this research.

Keywords: sentiment analysis, customer perception, market assessment, machine learning, YouTube, SMEs, natural language processing, BETO, VADER, TextBlob, CRISP-DM, Scrum, Python.

ÍNDICE

Capítulo 1: Introducción.....	15
1.1. Antecedentes	15
1.2. Definición del problema.....	18
1.2.1. Problema principal	18
1.2.2. Problemas secundarios	19
1.3. Objetivos	19
1.3.1. Objetivo General	19
1.3.2. Objetivos Específicos.....	19
1.4. Importancia.....	20
1.5. Justificación.....	22
1.6. Alcance.....	23
1.7. Organización de la tesis.....	24
Capítulo 2: Marco Teórico	26
2.1. Enfoques basados en Machine Learning.....	26
2.1.1. Aprendizaje Automático Supervisado.....	26
2.1.1.1. Árboles de decisión	26
2.1.1.2. SVM	27
2.1.1.3. Naive Bayes.....	28
2.1.2. Aprendizaje Automático No Supervisado	28
2.1.2.1. Clustering	28
2.1.3. Deep Learning	29
2.1.4. Redes Neuronales Artificiales.....	29
2.2. Enfoques basados en NLP	30
2.2.1. Enfoque basado en diccionarios.....	30
2.2.1.1. SentiWordNet	30
2.2.1.2. VADER.....	30
2.2.1.3. AFINN.....	30
2.2.2. Enfoque basado en corpus.....	31
2.2.2.1. TF-IDF	31
2.2.2.2. Word embeddings.....	31

2.3.	Metodología CRISP-DM.....	32
2.4.	Metodología Scrum.....	32
2.5.	Métricas de Evaluación.....	33
2.5.1.	AUC	33
2.5.2.	Accuracy.....	33
2.5.3.	F1.....	33
2.5.4.	Precision.....	34
2.5.5.	Recall.....	34
2.6.	Estilos de Arquitecturas de Software para el desarrollo de Plataformas Web	35
2.6.1.	Arquitectura en Capas (Layered)	35
2.6.2.	Arquitectura Cliente-Servidor	35
2.6.3.	Arquitectura de Microservicios.....	36
2.6.4.	Arquitectura Monolítica	36
2.6.5.	Arquitectura MVC (Modelo-Vista-Controlador).....	36
2.6.6.	Arquitectura Orientada a Servicios (SOA).....	37
Capítulo 3:	Estado del Arte	38
3.1.	Metodología	38
3.2.	Planificación.....	39
3.2.1.	Preguntas de investigación.....	39
3.2.2.	Bases de datos académicas.....	39
3.2.3.	Palabras clave.....	40
3.2.4.	Criterios de inclusión y exclusión.....	40
3.3.	Ejecución.....	41
3.4.	Resultados	43
3.5.	Análisis.....	47
3.5.1.	¿Qué herramientas y técnicas son las más efectivas para la recopilación y extracción de la información necesaria para el análisis de sentimientos?	47
3.5.2.	¿Cuáles son las técnicas más efectivas de preprocesamiento para preparar los datos en un formato adecuado para el análisis de sentimientos?	49
3.5.3.	¿Cuáles son las técnicas y algoritmos de clasificación de machine learning y NLP más efectivos y utilizados para el análisis de sentimientos?.....	51

3.5.4.	¿Qué métricas de validación son las más relevantes para evaluar la efectividad de los modelos de análisis de sentimientos?	53
3.5.5.	¿Qué criterios de diseño y desarrollo debe cumplir una plataforma web para garantizar que sea intuitiva, funcional y eficiente en la presentación de resultados de análisis de sentimientos a un usuario?	55
3.6.	Conclusión sobre hallazgos en la revisión sistemática.....	57
Capítulo 4: Diseño metodológico basado en CRISP-DM y SCRUM para el análisis de sentimientos.....		58
4.1.	Descripción del método.....	58
4.1.1.	Comprensión del negocio.....	61
4.1.2.	Comprensión de los datos	62
4.1.3.	Preparación de los datos	62
4.1.4.	Modelado.....	63
4.1.5.	Evaluación.....	64
4.1.6.	Despliegue.....	65
4.2.	Selección y justificación de la metodología.....	66
4.2.1.	Metodología para minería de datos	66
4.2.2.	Metodología para desarrollo ágil.....	67
4.3.	Esquema del desarrollo de la plataforma	69
Capítulo 5: Desarrollo de la Plataforma para el Análisis de Sentimientos en comentarios de YouTube.....		71
5.1.	Requerimientos para el desarrollo del sistema.....	71
5.1.1.	Hardware	71
5.1.2.	Software	71
5.2.	Arquitectura del sistema.....	72
5.3.	Descripción general del sistema.....	73
5.4.	Desarrollo de la solución.....	74
5.4.1.	Obtención y preprocesamiento de los datos.....	74
5.4.2.	Implementación de modelos preexistentes.....	78
5.4.2.1.	VADER.....	79
5.4.2.2.	BETO.....	80
5.4.2.3.	TextBlob	82

5.4.3.	Creación de la Plataforma Web.....	84
5.4.3.1.	Módulo de Extracción de Datos	85
5.4.3.2.	Módulo de Análisis de Sentimientos	86
5.4.3.3.	Módulo de Visualización de Resultados y Reportes	86
5.4.3.4.	Módulo de Información y Contacto	88
Capítulo 6:	Validación y Resultados	90
6.1.	Validación de Modelos de Análisis de Sentimientos	90
6.1.1.	Matriz de confusión.....	90
6.1.2.	Métricas de evaluación del modelo	92
6.1.3.	Proceso de validación de los modelos de análisis de sentimientos	94
6.1.4.	Resultado de validación de los modelos de análisis de sentimientos.....	96
6.2.	Validación de Plataforma Web	101
6.2.1.	Pruebas funcionales de validación	101
6.2.2.	Pruebas de usabilidad de validación.....	102
6.2.3.	Pruebas de rendimiento de validación.....	104
Capítulo 7:	Conclusiones y Trabajos Futuros	106
7.1.	Conclusión General	106
7.2.	Conclusiones Específicas	106
7.3.	Trabajos Futuros.....	107
7.4.	Recomendaciones.....	108
Referencias Bibliográficas		109

Lista de Figuras

Figura 1. Porcentaje de consumo de plataformas digitales para búsqueda de noticias.....	16
Figura 2. Número de usuarios de YouTube a nivel mundial	21
Figura 3. Proceso de revisión sistemática de la literatura para el estado del arte	42
Figura 4. Gráfica con la cantidad de artículos científicos por año de publicación	46
Figura 5. Gráfica con la cantidad de artículos científicos por banco de datos.....	47
Figura 6. Procesos de la Metodología CRISP-DM	59
Figura 7. Técnicas y enfoques para Análisis de Sentimientos	64
Figura 8. Esquema general del desarrollo de la plataforma web para análisis de sentimiento en comentarios de YouTube como herramienta de evaluación de mercado en PYMEs	69
Figura 9. Arquitectura del sistema propuesto	72
Figura 10. Código para la extracción del ID del video desde la URL	76
Figura 11. Código para la obtención de comentarios y respuestas del video de YouTube.....	76
Figura 12. Código para el preprocesamiento del dataset de comentarios de YouTube	78
Figura 13. Código para el análisis de sentimientos VADER	80
Figura 14. Código para el análisis de sentimientos BETO	82
Figura 15. Código para el análisis de sentimientos TextBlob.....	84
Figura 16. Interfaz de Inicio de la plataforma web	85
Figura 17. Interfaz de Módulo de Extracción de datos de la plataforma web	86
Figura 18. Interfaz de Reportes de Resultados de la plataforma web	87
Figura 19. Interfaz de Visualización de Resultados de la plataforma web	88
Figura 20. Interfaz de Información de la plataforma web.....	89
Figura 21. Interfaz de Contacto de la plataforma web	89
Figura 22. Esquema de Matriz de confusión de tres categorías	91

Lista de Tablas

Tabla 1. Tipos de anuncios en YouTube.....	17
Tabla 2. Criterios de inclusión y exclusión de estado del arte	40
Tabla 3. Cantidad de artículos seleccionados por base de datos	43
Tabla 4. Lista de los 21 artículos seleccionados para la investigación	43
Tabla 5. Herramientas y técnicas de obtención de datos para el análisis de sentimientos	48
Tabla 6. Métodos de preprocesamiento de datos para el análisis de sentimientos.....	49
Tabla 7. Técnicas de machine learning y NLP para el análisis de sentimientos.....	51
Tabla 8. Métricas de validación para modelos de análisis de sentimientos	54
Tabla 9. Criterios de desarrollo y diseño fundamentales de una plataforma web.....	56
Tabla 10. Integración de CRISP-DM y Scrum en el desarrollo de la plataforma web	60
Tabla 11. Comparativa de metodologías KDD, CRISP-DM y SEMMA.....	66
Tabla 12. Comparativa de metodología Scrum y Kanban.....	68
Tabla 13. Atributos del dataset generado a partir de la API de YouTube.....	75
Tabla 14. Lista de 10 videos utilizados para la validación del modelo.....	94
Tabla 15. Matriz de Confusión Consolidada para la Categoría: Reseñas de Productos o Servicios Populares	97
Tabla 16. Matriz de Confusión Consolidada para la Categoría: Campañas publicitarias de marcas	97
Tabla 17. Matriz de Confusión Consolidada para la Categoría: Reseñas de negocios o pequeños locales.....	98
Tabla 18. Evaluación de métricas de validación en modelos de Análisis de Sentimientos con casos de prueba predefinidos.....	99
Tabla 19. Validación funcional de los endpoints de la plataforma	102
Tabla 20. Resultado de evaluación subjetiva de usabilidad de la plataforma web	103
Tabla 21. Evaluación del rendimiento por modelo en extracción y análisis de comentarios.....	105

Capítulo 1: Introducción

1.1. Antecedentes

La era digital, también conocida como sociedad de la información, es un período histórico que se inicia con la creación de dispositivos digitales y se extiende hasta la actualidad. Esta época se caracteriza por el uso de las tecnologías de información, la expansión de los medios de comunicación, la explosión de las redes sociales y la incorporación de nuevas formas de cultura digital. Esta revolución ha provocado que la implementación de las tecnologías de la información redefina los medios de comunicación tradicionales.

Un ejemplo de este cambio es la investigación realizada por el Reuters Institute for the Study of Journalism, en colaboración con la Universidad de Oxford, en la que se analizan las plataformas más utilizadas para la búsqueda de noticias e información en algunos países latinoamericanos, como se puede observar en la Figura 1.

La digitalización conlleva la recopilación de una cantidad significativa de información proveniente de comentarios y respuestas de usuarios en distintos medios en línea como videos, publicaciones en redes sociales, páginas web, blogs y cursos digitales. Esta información refleja opiniones, percepciones y experiencias de los usuarios, cuyo análisis permite a las empresas obtener un panorama más preciso de las necesidades actuales del mercado. El análisis adecuado de datos no estructurados, como comentarios de redes sociales, encuestas o chats, ofrece ventajas competitivas al revelar tendencias de consumo, fallas en productos o intereses de los clientes. (Ramaswamy & DeClerck, 2018)

En este contexto, el análisis de sentimientos se presenta como una herramienta clave para extraer información valiosa de los datos textuales online. Permite procesar el contenido escrito por los usuarios y evaluar las emociones expresadas en lenguaje natural.

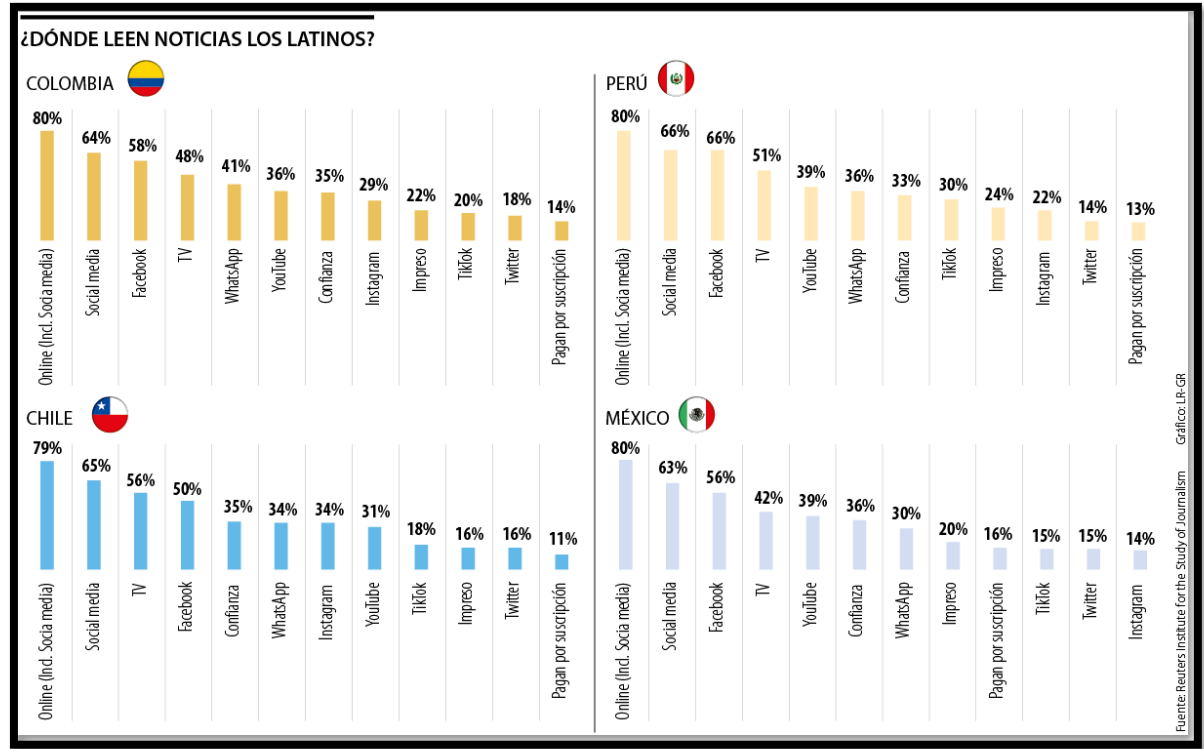


Figura 1. Porcentaje de consumo de plataformas digitales para búsqueda de noticias

Fuente: Reuters Institute for the Study of Journalism, University of Oxford (2024), en el informe *Digital News Report 2024*

Entre las diversas plataformas digitales donde los usuarios expresan sus opiniones e interactúan activamente, YouTube destaca como una de las más relevantes. Creada en 2005, sigue vigente con millones de usuarios activos mensuales, constituyéndose como una de las redes sociales más antiguas y utilizadas para difundir contenido por parte de usuarios y empresas.

En este sentido, se considera la plataforma YouTube como el foco principal de la presente investigación por su uso extendido entre marcas que buscan alcanzar nuevos clientes, fidelizar

a los clientes existentes y difundir campañas de marketing. Investigaciones recientes respaldan el uso de YouTube frente a otras plataformas con fines informativos. De Oliveira Júnior et al. (2024) realizaron un estudio comparativo sobre videos del ámbito de la salud y determinaron que los videos de YouTube ofrecían una calidad informativa superior a los de TikTok, gracias a su mayor profundidad, duración y respaldo de instituciones reconocidas. Asimismo, Nguyen y Diederich (2023) evidenciaron que el formato breve y distractor de TikTok afecta negativamente la memoria prospectiva, es decir, la capacidad cognitiva para recordar intenciones o tareas futuras.

Según un artículo publicado en la página oficial de Hootsuite, YouTube es el segundo sitio web más visitado del mundo, con más de 3 000 millones de usuarios mensuales, lo que lo convierte en un espacio idóneo para analizar la percepción de los usuarios frente al contenido en línea. (Cooper, 2022) Además, YouTube ofrece herramientas estratégicas para las empresas, como YouTube Shorts, anuncios publicitarios en los videos, tarjetas interactivas y pantallas finales que optimizan la visibilidad del contenido. La Tabla 1 muestra los principales tipos de anuncios en YouTube y sus características.

Tabla 1. Tipos de anuncios en YouTube

Fuente: Elaboración propia en base a información Hootsuite (Cooper, 2022)

Tipo de Anuncio	Característica Principal
Anuncios de video con opción de omitir	Videos publicitarios que se muestran antes o durante la reproducción del contenido (pre-roll o mid-roll). Los usuarios pueden omitirlos tras 5 segundos. Se recomienda que no excedan los 3 minutos para mantener la atención.
Anuncios de video sin opción de omitir	Videos que se reproducen completamente sin posibilidad de omitir. Son ideales para campañas que buscan un fuerte posicionamiento de marca y

	confían en su capacidad creativa para captar la atención durante unos 15 segundos.
Anuncios in-feed	Similares a los anuncios en los resultados de búsqueda de Google. Aparecen junto a contenido relacionado, atrayendo visualizaciones si resultan más relevantes que otros resultados.
Anuncios display	Ideales para quienes no cuentan con contenido en video. Incluyen banners en la parte lateral derecha de la página y elementos superpuestos sobre los videos, con imagen, texto, llamado a la acción y enlace al sitio web.

En Perú, las pequeñas y medianas empresas (PYMEs) desempeñan un papel crucial en la economía nacional, generando una parte significativa del empleo. De acuerdo con un estudio del Ministerio de Producción (PRODUCE), al cierre de 2023 se registraron 2 259 534 empresas formales registradas en la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT). De esta cifra el 99,2% correspondía a micro y pequeñas empresas (MYPE), lo que confirma su relevancia en el tejido empresarial. (Programa Nacional “Tu Empresa”, 2024)

1.2. Definición del problema

1.2.1. Problema principal

¿Cómo pueden las pequeñas y medianas empresas (PYMEs) del Perú contar con herramientas automatizadas y accesibles que les permitan recopilar, procesar, analizar y presentar eficientemente las reseñas y preferencias de los clientes sobre un producto o servicio en idioma español, a fin de mejorar su crecimiento y competitividad en el mercado actual?

1.2.2. Problemas secundarios

- ¿Cómo se puede fortalecer la capacidad de toma de decisiones empresariales de las PYMEs mediante el acceso a datos precisos y relevantes sobre la experiencia de los usuarios con un producto?
- ¿Cómo implementar métodos efectivos y accesibles para extraer y recopilar información relevante desde plataformas digitales como YouTube?
- ¿De qué forma promover la adopción de herramientas basadas en inteligencia artificial que permitan visualizar e identificar tendencias y necesidades del cliente en plataformas digitales para optimizar las estrategias de negocio?

1.3. Objetivos

1.3.1. Objetivo General

Diseñar y desarrollar una plataforma web para el análisis de sentimientos de comentarios en idioma español de videos de YouTube, que permita a las PYMEs obtener información sobre la percepción de sus clientes mediante el uso de enfoques y algoritmos de Machine Learning y Procesamiento de Lenguaje Natural (NLP).

1.3.2. Objetivos Específicos

- Implementar un método eficiente para la extracción y recopilación de comentarios en español de videos de YouTube para su posterior análisis.
- Aplicar técnicas de preprocesamiento efectivas que permitan limpiar y preparar los datos de los comentarios en español, asegurando su calidad para el análisis de sentimientos.

- Identificar y seleccionar modelos basados en métodos adecuados de Machine Learning o NLP para realizar el análisis de sentimientos de comentarios en español.
- Definir y emplear métricas de validación adecuadas para evaluar la eficiencia y precisión de los modelos de análisis de sentimiento implementados.
- Diseñar y desarrollar una plataforma web intuitiva, funcional y responsiva que permita a los usuarios visualizar los resultados del análisis de sentimientos de manera comprensible, mediante información detallada y gráficos ilustrativos.

1.4.Importancia

En el entorno actual, las opiniones y comentarios de los clientes constituyen elementos esenciales de la comunicación digital, ya que aportan información valiosa para la formulación de estrategias empresariales. Dentro del análisis de mercado, uno de los factores más determinantes es la percepción del consumidor, pues influye directamente en las decisiones de compra, la fidelidad hacia la marca y la posición competitiva de la empresa mediante el ajuste de sus estrategias comerciales. Como señala Porlles et al. (2006), la evaluación del mercado considera no solo aspectos cuantitativos, sino también cualitativos, tales como las preferencias del consumidor, las barreras percibidas por los usuarios y los hábitos de consumo, los cuales influyen en las decisiones comerciales a largo plazo.

En este contexto, YouTube se consolida como una de las plataformas digitales con mayor alcance global, al registrar miles de millones de usuarios activos mensualmente y un elevado tiempo promedio de visualización diaria. Este volumen y la constante interacción convierten a YouTube en un canal idóneo para recopilar y analizar opiniones de los consumidores. La página Statista, reconocida por ofrecer datos estadísticos actualizados sobre diversas temáticas de

interés global, presenta un artículo en el que se proyecta un crecimiento sostenido de usuarios de YouTube hasta el año 2028 a nivel mundial, lo cual evidencia su posicionamiento como una de las plataformas más influyentes a escala global. (Orús, 2024) Este crecimiento proyectado se muestra en la Figura 2.

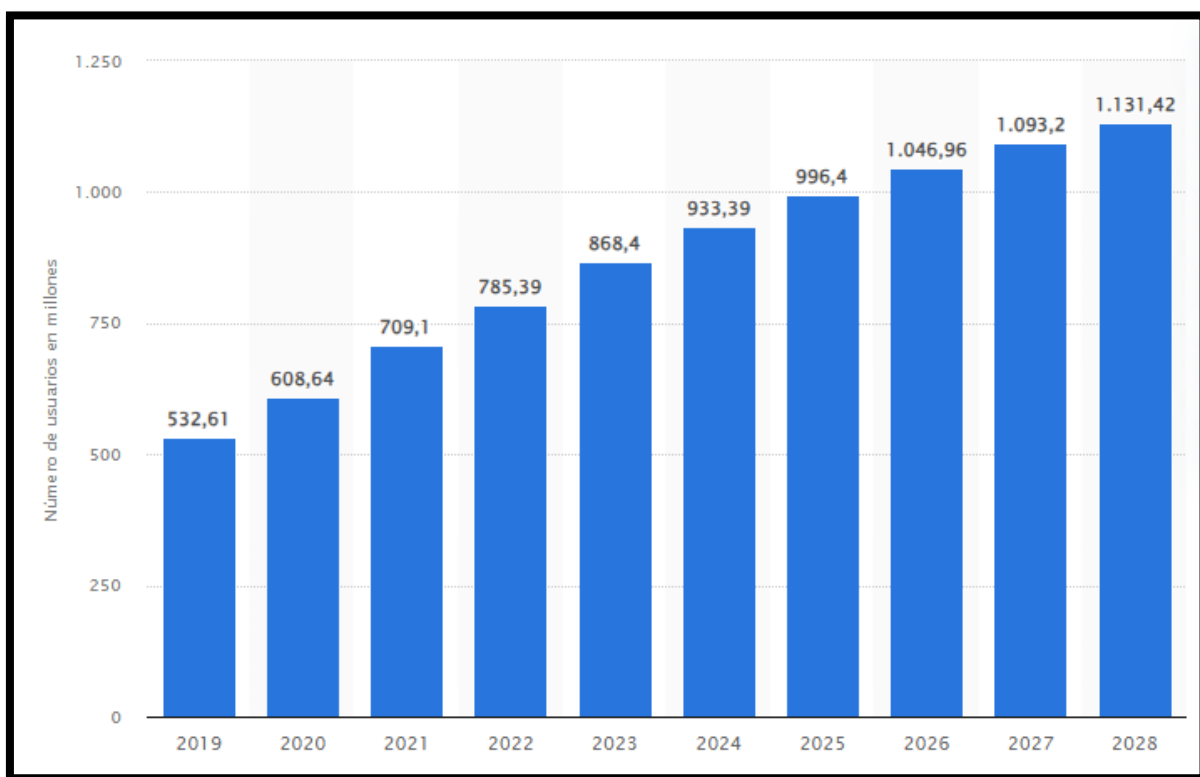


Figura 2. Número de usuarios de YouTube a nivel mundial

Fuente: Artículo de la página web Statista (Orús, 2024)

El análisis de sentimientos mediante el uso de algoritmos y modelos permite comprender y clasificar automáticamente las opiniones de los consumidores en categorías como positivas, negativas o neutras. Esta clasificación facilita la interpretación masiva de datos y otorga a las empresas una ventaja competitiva al transformar la opinión del cliente en información accionable.

Esta investigación busca generar un impacto significativo en el sector tecnológico, promoviendo el acceso a tecnologías de análisis de datos mediante Machine Learning y NLP por parte de PYMEs con recursos limitados. Si bien existen plataformas dedicadas a la gestión y seguimiento de opiniones y reseñas de clientes en línea, la mayoría están orientadas al mercado angloparlante, lo que genera una brecha tecnológica para las PYMEs hispanohablantes. Entre estos servicios se encuentran Broadly, BirdEye, GatherUp, ReviewTrackers y Trustpilot. En síntesis, la relevancia de esta investigación radica en ofrecer acceso a tecnologías de análisis de la percepción del cliente, mediante Machine Learning y Procesamiento de Lenguaje Natural en el idioma español, con el propósito de apoyar a las PYMEs y contribuir al desarrollo de soluciones accesibles y adaptadas al contexto empresarial hispanohablante. Esta necesidad se ve acentuada por la predominancia de soluciones desarrolladas para el idioma inglés, que no consideran las particularidades lingüísticas y semánticas del español, lo que puede generar interpretaciones inexactas y limitar la utilidad de los resultados.

1.5. Justificación

A pesar de que las PYMEs representan un porcentaje significativo en diversos sectores económicos del país, muchas de ellas aún realizan el análisis de mercado, especialmente el relacionado con la percepción del cliente, de forma manual o sin apoyo tecnológico especializado. De acuerdo con Hulla et al. (2021), numerosas PYMEs operan sin una estrategia de digitalización, lo que las lleva a depender de métodos tradicionales como encuestas, entrevistas o la observación directa del comportamiento del cliente. Estas prácticas, si bien son útiles, resultan limitadas en eficiencia y alcance, dificultando su adaptación a la transformación

digital y el aprovechamiento de tecnologías de información que permitan una mejor visualización y análisis de datos.

Desde el enfoque tecnológico y académico, gran parte de los estudios recientes sobre análisis de sentimientos aplicados a comentarios en redes sociales se han centrado en contenidos en idioma inglés. Esto se debe a que la mayoría de los modelos y recursos han sido desarrollados originalmente para dicho idioma, lo cual representa una limitación significativa al aplicarlos directamente en contextos hispanohablantes (Korayem et al., 2016).

En ese sentido, la presente investigación busca aportar conocimiento en la aplicación de técnicas de análisis de sentimientos en comentarios en español publicados en redes sociales, un campo aún poco explorado que requiere soluciones adaptadas tanto a las particularidades lingüísticas del idioma como al contexto sociocultural latinoamericano.

1.6. Alcance

Esta investigación propone el desarrollo de una plataforma web que permita al usuario ingresar la URL de un video de YouTube, a partir de la cual se extraerán y procesarán los comentarios en idioma español mediante técnicas de Machine Learning y NLP. El objetivo es determinar la recepción del video por parte de los usuarios a través de la clasificación automática del sentimiento en tres categorías: positivo, negativo o neutro.

Además, la plataforma proporcionará resultados mediante gráficos e indicadores clave que faciliten la interpretación de la percepción del cliente, brindando así apoyo estratégico a las pequeñas y medianas empresas (PYMEs) en sus procesos de toma de decisiones comerciales. Cabe precisar que el enfoque de análisis de mercado abordado en esta investigación se restringe

al estudio de la percepción del cliente mediante la interpretación automatizada de comentarios en español publicados en YouTube.

El análisis se limitará a comentarios en idioma español y los resultados estarán orientados principalmente a proporcionar información relevante para el contexto de las PYMEs en el Perú.

Asimismo, la investigación se centra únicamente en la plataforma de YouTube, sin contemplar el análisis de opiniones en otras redes sociales.

1.7. Organización de la tesis

La presente tesis se estructura en 7 capítulos, los cuales se describen a continuación:

En el capítulo 2, se presenta el Marco Teórico donde se revisan los conceptos en los que se ha sustentado la investigación.

En el capítulo 3, se presenta el Estado del Arte donde se revisan trabajos previos y se define la metodología implementada para la revisión de la literatura que consta de la planificación de la investigación, el desarrollo de la investigación, los resultados y el análisis (preguntas de investigación).

En el capítulo 4, se presenta el Diseño Metodológico basado en CRISP-DM y Scrum como guía para el desarrollo de la solución, junto con su justificación frente a otras metodologías vigentes y un esquema orientado por dichas metodologías que describe el proceso de construcción de la plataforma web.

En el capítulo 5 se presenta el Desarrollo de Plataforma para el Análisis de Sentimiento en comentarios en idioma español de YouTube donde se describirá el desarrollo del sistema, la arquitectura del sistema, los requerimientos de funcionamiento y la implementación de la solución definida.

En el capítulo 6 se presenta el proceso de Validación y Resultados de los modelos de análisis de sentimientos y de la plataforma web. Se verifica el desempeño de los modelos mediante métricas de evaluación y se realizan pruebas funcionales, de usabilidad y de rendimiento para garantizar el correcto funcionamiento de la plataforma, además se discuten los resultados obtenidos en cada evaluación.

En el capítulo 7 se presentan las Conclusiones, Trabajos Futuros y Recomendaciones, donde se indicarán las principales conclusiones obtenidas luego de la investigación, se brindarán recomendaciones y se sugerirán puntos a abordar en trabajos futuros.

Finalmente, se incluyen las referencias bibliográficas correspondientes a las fuentes citadas.

Capítulo 2: Marco Teórico

En este capítulo se presentan los fundamentos teóricos que sustentan la investigación, donde se definen los conceptos, métricas, metodologías y algoritmos clave para facilitar la comprensión de los capítulos posteriores. Estos conceptos serán esenciales para comprender el diseño, desarrollo y validación de la plataforma propuesta.

2.1. Enfoques basados en Machine Learning

2.1.1. Aprendizaje Automático Supervisado

Según Oracle (2024), los algoritmos de aprendizaje automático supervisados son los más utilizados. En este enfoque, el científico de datos actúa como guía, enseñando al algoritmo las conclusiones que debe obtener. El algoritmo se entrena utilizando un conjunto de datos etiquetados con resultados predefinidos.

2.1.1.1. Árboles de decisión

Los algoritmos de árboles de decisión se utilizan principalmente para problemas de clasificación y el modelado predictivo. Al clasificarlos por sus valores, dividen los atributos en dos o más clases. En cada árbol hay nodos y ramas. Cada nodo representa un atributo del grupo y cada valor posible está representado por una rama. (Dankar & Gupta, 2021)

Similar a un diagrama de flujo, un árbol de decisión comienza con un nodo raíz, que formula una pregunta concreta a los datos y luego los envía por una rama en función de la respuesta. Cada una de estas ramas conduce a un nodo interno, que a su vez formula otra pregunta a los datos antes de dirigirlos hacia otra rama en función de

la respuesta. Esto continúa hasta que los datos llegan a un nodo final que no se ramifica más.

Random Forest es un algoritmo basado en múltiples árboles de decisión que operan en conjunto. Cada árbol realiza una predicción individual y luego se combinan las decisiones para determinar la clase final más representativa. Este modelo se destaca por su alta precisión en tareas de clasificación, siendo especialmente eficaz al trabajar con grandes volúmenes de datos y gestionar un elevado número de variables de entrada.

2.1.1.2. SVM

Las máquinas de soporte vectorial son algoritmos de aprendizaje supervisado que permiten clasificar datos en dos grupos mediante una separación lineal. La idea principal del SVM es la fijación de un hiperplano que separa datos de entrada en subgrupos. Este debe estar ubicado de tal manera que se maximice la distancia entre él y los valores de entrada más cercanos, estos valores más cercanos son los llamados vectores de soporte.

Los vectores de soporte son un subconjunto de los datos que son los puntos más próximos de las clases y definen la posición del hiperplano ideal. Los otros puntos de datos no se consideran al determinar ese hiperplano. Los vectores de soporte corresponden a un grupo de datos, caracterizados por ser los elementos más cercanos a la línea o plano de separación entre clases. Estos puntos son fundamentales para definir la ubicación óptima del hiperplano, mientras que el resto de los datos no influyen directamente en ella. (Aguilar & Camargo, 2021)

2.1.1.3. Naive Bayes

El clasificador Naive Bayes es un conjunto de algoritmos de aprendizaje supervisado que se utilizan para crear modelos predictivos de categorización binaria o múltiple. Basado en el Teorema de Bayes, Naive Bayes opera con probabilidades condicionales, que son independientes entre sí, pero indican la probabilidad de una clasificación basada en sus factores combinados. Es ideal para tareas de clasificación rápida como el procesamiento de texto y análisis de sentimientos.

2.1.2. Aprendizaje Automático No Supervisado

Según Oracle (2024), el aprendizaje automático no supervisado utiliza un enfoque más independiente en el que un ordenador aprende a identificar procesos y patrones complejos sin la guía cercana y constante de un ser humano. El aprendizaje automático no supervisado implica un entrenamiento basado en datos que no tienen ni etiquetas ni un resultado específico definido.

2.1.2.1. Clustering

Es un método de aprendizaje no supervisado donde se separan los objetos en varios lotes o grupos basados en su similitud. Los objetos dentro de un mismo clúster comparten más similitudes entre sí que con los objetos pertenecientes a otros clústeres. (Dankar & Gupta, 2021)

K-means es un algoritmo no supervisado utilizado para resolver el problema de clustering, que permite agrupar datos en diferentes clústeres basados en su similitud. Se emplea comúnmente en tareas de segmentación de datos y análisis exploratorio. K-means utiliza la proximidad de un resultado a un conglomerado de puntos de datos para identificarlo. Cada uno de los conglomerados está definido por un

centroide, un punto central real o imaginario del conglomerado. K-means es útil en grandes conjuntos de datos, especialmente para la agrupación, aunque puede fallar cuando maneja valores atípicos.

2.1.3. Deep Learning

El aprendizaje profundo o deep learning es una rama del aprendizaje automático que emplea redes neuronales profundas con múltiples capas para procesar grandes volúmenes de información. Estas redes, inspiradas en el funcionamiento del cerebro humano, son capaces de identificar y representar patrones complejos a partir de datos masivos. Utiliza capas de redes neuronales artificiales que computan con representaciones continuas de números reales, reflejando un poco cómo las neuronas se organizan jerárquicamente en los cerebros humanos. (Rodríguez et al., 2024)

2.1.4. Redes Neuronales Artificiales

Una red neuronal es un tipo de sistema informático que simula el funcionamiento del cerebro humano y que constituye una base fundamental en el desarrollo de la inteligencia artificial. Se compone de neuronas artificiales que tienen conexiones ponderadas entre las unidades. En síntesis, las redes neuronales consisten en capas de neuronas interconectadas que aprenden patrones complejos a través de múltiples iteraciones de entrenamiento.

2.2.Enfoques basados en NLP

2.2.1. Enfoque basado en diccionarios

Según Alaminos (2023), los enfoques basados en léxico para el análisis de sentimiento utilizan listas de palabras predefinidas, cada una asociada a emociones o sentimientos específicos. Estos enfoques utilizan diccionarios que permiten identificar y cuantificar la polaridad expresada en un texto. Ejemplos de enfoques léxicos para el análisis de sentimientos incluyen SentiWordNet, VADER y AFINN.

2.2.1.1. SentiWordNet

Es un léxico de sentimiento basado en WordNet que contiene conjuntos de sinónimos llamados synsets. Cada synset en SentiWordNet tiene asociaciones de sentimiento positivas, negativas y objetivas. Es ideal para analizar el sentimiento en textos más formales y en contextos donde la desambiguación semántica es crucial (Alaminos, 2023)

2.2.1.2. VADER

Es un léxico de sentimiento y un modelo de análisis de sentimiento basado en reglas. Es ideal para analizar el sentimiento en textos informales, como los encontrados en redes sociales, pues considera la intensificación y la negación de las palabras. Sus siglas, VADER, significan Valence Aware Dictionary and Sentiment Reasoner. (Alaminos, 2023)

2.2.1.3. AFINN

Es un léxico de sentimiento que consta de una lista de palabras en inglés asociadas con puntuaciones de sentimiento que oscilan entre -5 (muy negativo) y +5 (muy

positivo). Es ideal para analizar el sentimiento en redes sociales y mensajes cortos, ya que incluye palabras coloquiales y expresiones informales. (Alaminos, 2023)

2.2.2. Enfoque basado en corpus

Según Torrijos y Goñi (2022), el enfoque basado en corpus en el procesamiento del lenguaje natural (NLP) implica el uso de colecciones extensas de textos anotados, para desarrollar modelos de aprendizaje automático. Este enfoque se centra en tres fases principales: la anotación del corpus original, la conversión del corpus anotado al formato adecuado para el entrenamiento y la obtención del dataset final para el entrenamiento del modelo. Para transformar el texto en un formato procesable, se utilizan diversas técnicas de representación del texto:

2.2.2.1. TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) es una medida ponderada que se utiliza para evaluar la importancia de una palabra en un conjunto de documentos, teniendo en cuenta la frecuencia de la palabra en los documentos y la rareza de la palabra en todo el conjunto de documentos. (Alaminos, 2023)

2.2.2.2. Word embeddings

Los Word embeddings son representaciones vectoriales de palabras en un espacio de alta dimensión, que capturan relaciones semánticas y sintácticas entre palabras basadas en su contexto de co-ocurrencia en un corpus de texto. (Alaminos, 2023)

Word2Vec y GloVe son ejemplos de técnicas basadas en corpus que generan estas representaciones vectoriales.

2.3. Metodología CRISP-DM

CRISP-DM (Cross Industry Standard Process for Data Mining) es una metodología utilizada en proyectos de análisis de datos que proporciona un marco estructurado en seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue.

De acuerdo con Chapman et al. (2000), la metodología CRISP-DM se aplica especialmente en proyectos donde se requiere transformar datos en conocimiento útil para la toma de decisiones. Sus principales beneficios incluyen la adaptabilidad a diferentes contextos de estudio, el mantenimiento de un buen flujo de trabajo, la facilidad de documentación de cada etapa y su naturaleza iterativa que permite retroceder a fases previas ya sea para corregir errores, mejorar el modelo o replantear objetivos.

2.4. Metodología Scrum

Scrum es un marco ágil de trabajo orientado al desarrollo iterativo e incremental de productos mediante el uso de ciclos cortos de trabajo llamados “sprints”. Según lo planteado por Schwaber y Sutherland (2020), Scrum se aplica en contextos donde los requerimientos cambian constantemente y se necesita un enfoque que permita a los equipos adaptarse ante nuevas necesidades, siendo utilizado en proyectos de desarrollo de software como aplicaciones web y productos digitales. Entre sus principales beneficios destacan la retroalimentación constante, la mejora continua y la visibilidad del avance del proyecto.

2.5. Métricas de Evaluación

2.5.1. AUC

Área bajo la curva (AUC por sus siglas en inglés) se calcula usando el área bajo la curva ROC. Cuanto mayor es el área, más preciso es formalmente el predictor. La fórmula para calcular el AUC es representada por la siguiente ecuación:

$$AUC = \int_0^1 f(x) dx$$

Donde $f(x)$ representa la función de la curva característica de funcionamiento del receptor (ROC por sus siglas en inglés). (Morales et al., 2021)

2.5.2. Accuracy

Accuracy es el grado de cercanía al valor verdadero. Se refiere a una medición con resultados tanto verdaderos como consistentes. La fórmula para calcular el Accuracy es representada por la siguiente ecuación:

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n}$$

Donde t_p representa un valor verdadero-positivo, t_n un valor verdadero-negativo, f_p un valor falso-positivo y f_n un valor falso-negativo. (Morales et al., 2021)

2.5.3. F1

F1 es una métrica que se calcula a partir de la precisión y el recall de la prueba que se lleva a cabo. En pocas palabras, F1 es la media armónica de la precisión y el recall, que se muestra en la ecuación:

$$F1 = \frac{t_p}{t_p + \frac{1}{2}(f_p + f_n)}$$

Donde t_p representa un valor verdadero-positivo, f_p un valor falso-positivo y f_n un valor falso-negativo. (Morales et al., 2021)

2.5.4. Precision

Precision es una métrica de rendimiento que se aplica en datos recuperados de una colección, corpus o espacio muestral. También se le conoce como valor predictivo positivo, que es una fracción de instancias relevantes entre las instancias recuperadas, tal como se muestra en la ecuación:

$$Precision = \frac{t_p}{t_p + f_p}$$

Donde t_p equivale a un valor verdadero-positivo y f_p a un valor falso-positivo. (Morales et al., 2021)

2.5.5. Recall

Recall, también conocido como sensibilidad, es una fracción de instancias relevantes que se lograron recuperar. La fórmula para calcular el Recall es representada por la siguiente ecuación:

$$Recall = \frac{t_p}{t_p + f_n}$$

Donde t_p representa un valor verdadero-positivo y f_n representa un valor falso-negativo. (Morales et al., 2021)

2.6. Estilos de Arquitecturas de Software para el desarrollo de Plataformas Web

2.6.1. Arquitectura en Capas (Layered)

Como menciona Richards (2015), los componentes del patrón de arquitectura en capas se organizan en capas horizontales, donde cada una desempeña una función específica dentro de la aplicación (por ejemplo, lógica de presentación o lógica empresarial). Aunque el patrón de arquitectura en capas no especifica el número ni los tipos de capas que deben existir, la mayoría de las arquitecturas en capas constan de cuatro capas estándar: presentación, negocio, persistencia y base de datos. En algunos casos, la capa de negocio y la capa de persistencia se combinan en una única capa de negocio. Así, las aplicaciones más pequeñas pueden tener solo tres capas, mientras que las aplicaciones de negocio más grandes y complejas pueden contener cinco o más capas.

2.6.2. Arquitectura Cliente-Servidor

Según lo planteado por Lizama et al. (2016), este estilo de arquitectura se divide en dos partes principales:

- **Cliente:** Se define al cliente como un software que inicia la comunicación dentro de un entorno cliente-servidor, solicitando servicios o recursos y esperando una respuesta desde el servidor. Los clientes generalmente son aplicaciones que interactúan con el usuario, como navegadores web, aplicaciones móviles o software de escritorio.
- **Servidor:** El servidor es un programa o sistema que ofrece un servicio accesible a través de una red. Acepta solicitudes de los clientes, procesa las peticiones y

devuelve los resultados correspondientes. Los servidores pueden manejar múltiples conexiones simultáneamente y proporcionar diversos servicios.

2.6.3. Arquitectura de Microservicios

Es un enfoque para el desarrollo de aplicaciones que se compone de un conjunto de pequeños servicios, cada uno ejecutándose en su propio proceso y comunicándose mediante mecanismos ligeros. Estos servicios están contruidos en torno a las capacidades del negocio, con independencia en el despliegue e implementación totalmente automatizada. La arquitectura de microservicios promueve el desarrollo y despliegue de aplicaciones compuestas por unidades independientes, autónomas, modulares y autocontenidas. (López & Maya, 2017)

2.6.4. Arquitectura Monolítica

Según Guimarey (2020), un sistema monolítico es aquel en el que todos los módulos relevantes se encuentran empaquetados en una única pieza y todas las funcionalidades deben desplegarse juntas. En este enfoque, la vista, la lógica de negocio y la capa de acceso a los datos son componentes que residen juntos en un mismo sistema. Al inicio del desarrollo de un producto y de un equipo, los monolitos son ideales ya que permiten evitar el rediseño y simplifican el mantenimiento.

2.6.5. Arquitectura MVC (Modelo-Vista-Controlador)

De acuerdo con Pantoja (2004), el patrón MVC divide una aplicación web en tres módulos:

- **Modelo:** Es la parte del sistema que organiza y representa los datos del entorno real mediante clases que permiten su tratamiento dentro de la aplicación. El modelo es responsable de gestionar los datos y contiene la lógica de negocio de la aplicación

para recuperar, almacenar y actualizar los datos. El modelo desconoce la existencia de las vistas y del controlador.

- Vista: Es un conjunto de clases responsables de presentar al usuario los datos provenientes del modelo. Se pueden tener múltiples vistas vinculadas a un mismo modelo, ya que su función es representar visualmente la información sin modificarla.
- Controlador: El controlador actúa como intermediario que gestiona el flujo de la aplicación en respuesta a eventos externos, como entradas del usuario o interacciones con la interfaz. Su rol consiste en actualizar el modelo o gestionar la visualización al abrir o cerrar vistas. Aunque tiene acceso tanto al modelo como a las vistas, estos últimos no mantienen conocimiento directo del controlador.

2.6.6. Arquitectura Orientada a Servicios (SOA)

Como establece Erl (2008), SOA representa un modelo arquitectónico cuyo objetivo es mejorar la agilidad y la rentabilidad de una empresa, al tiempo que reduce la carga de TI. SOA se caracteriza por la introducción de nuevas tecnologías y plataformas que apoyan específicamente la creación, ejecución y evolución de soluciones orientadas a servicios, siendo este el componente principal de la lógica de la solución.

SOA define una manera de hacer que los componentes de software sean reutilizables a través de interfaces de servicio. Los servicios utilizan estándares de interfaz comunes y un patrón arquitectónico que permite su incorporación rápida a nuevas aplicaciones. De este modo, se evita que el desarrollador de aplicaciones deba volver a desarrollar, duplicar funcionalidades existentes o lidiar con la interoperabilidad entre funciones ya implementadas. (IBM, 2023)

Capítulo 3: Estado del Arte

Este capítulo presenta una revisión de la literatura relacionada con los métodos y fundamentos para la aplicación del análisis de sentimientos. En primer lugar, se formularán preguntas clave que guiarán el rumbo de la investigación. El objetivo es responder a estas preguntas mediante la búsqueda sistemática y el análisis comparativo de artículos científicos recientes. Finalmente, se desarrollará una taxonomía que permita comparar los distintos estudios e identificar vacíos de conocimiento en los que se pueda indagar y proponer una solución innovadora.

3.1. Metodología

Gómez Vargas et al. (2015) plantean una metodología para la elaboración del estado del arte que consta de las siguientes fases: planificación, ejecución, resultados y análisis.

- Planificación: Se formulan las preguntas de investigación y se definen los criterios de inclusión y exclusión para la selección de artículos científicos.
- Ejecución: Se buscan y seleccionan los artículos científicos de acuerdo con los criterios definidos en la fase previa, con el fin de asegurar una revisión sistemática de la literatura.
- Resultados: Se presentan los artículos más relevantes obtenidos tras la fase de ejecución.
- Análisis: Se responden las preguntas de investigación con el fin de identificar las principales contribuciones de los artículos revisados y su relación con la presente investigación, evidenciando así la relevancia del estudio al integrar conocimientos significativos.

3.2. Planificación

En esta fase se han definido las preguntas de investigación, las bases de datos bibliográficas, las palabras clave y los criterios de inclusión y exclusión para seleccionar los artículos relevantes a la investigación.

3.2.1. Preguntas de investigación

Para lograr el objetivo de la presente investigación, se plantean las siguientes preguntas:

Q1: ¿Qué herramientas y técnicas son las más efectivas para la recopilación y extracción de la información necesaria para el análisis de sentimientos?

Q2: ¿Cuáles son las técnicas más efectivas de preprocesamiento para preparar los datos en un formato adecuado para el análisis de sentimientos?

Q3: ¿Cuáles son las técnicas y algoritmos de clasificación de machine learning y NLP más efectivos y utilizados para el análisis de sentimientos?

Q4: ¿Qué métricas de validación son las más relevantes para evaluar la efectividad de los modelos de análisis de sentimientos?

Q5: ¿Qué criterios de diseño y desarrollo debe cumplir una plataforma web para garantizar que sea intuitiva, funcional y eficiente en la presentación de resultados de análisis de sentimientos a un usuario?

3.2.2. Bases de datos académicas

Para la búsqueda de los artículos científicos, se han considerado las siguientes fuentes:

Science Direct, IEEE Xplore, ACM Digital Library y Springer.

3.2.3. Palabras clave

Para la búsqueda de los artículos científicos se han considerado palabras clave tanto en inglés como en español. Las principales palabras clave utilizadas fueron: “análisis de sentimientos”, “sentiment analysis”, “opinion mining”, “procesamiento de lenguaje natural”, “YouTube”, “Twitter”, “machine learning”, “web scraping”, “web platform”, “plataforma web”, “comments”, “customer perception”, “PYMES”.

A partir de estas palabras clave, se formularon diversas cadenas de búsqueda combinadas mediante operadores booleanos (AND, OR), permitiendo identificar artículos relevantes para el desarrollo del estado del arte de esta investigación. Algunos ejemplos son:

- “sentiment analysis” AND “machine learning”
- “plataforma web” AND “análisis de sentimientos”
- “opinion mining” OR “sentiment analysis”
- “customer perception” AND “market evaluation”

3.2.4. Criterios de inclusión y exclusión

Se establecieron los siguientes criterios de inclusión y exclusión para la selección de los artículos científicos que aportarán a los objetivos de la presente investigación:

Tabla 2. Criterios de inclusión y exclusión de estado del arte

Fuente: Elaboración propia

Criterios de inclusión	Criterios de exclusión
Estudios o artículos que presenten propuestas (modelos, técnicas o herramientas) que sostengan la investigación con fundamentos en análisis de sentimiento	Estudios o artículos cuyo tema de investigación no se centre en análisis de sentimiento, procesamiento de lenguaje natural, machine learning o NLP.

utilizando técnicas de machine learning o NLP.	
Estudios o artículos que utilicen las reseñas o comentarios recopilados de una red social para verificar las diversas técnicas de obtención de un dataset.	Estudios o artículos cuyo resumen indique que no tiene relación con la investigación propuesta o tiene un enfoque diferente.
Estudios o artículos cuyo idioma de estudio esté en inglés o en español, con el fin de abarcar más información y aplicaciones.	Estudios o artículos cuyo idioma de estudio no sea ni en inglés ni español.
Estudios o artículos recientes, para tener en cuenta las tendencias actuales, publicados durante el período de tiempo: 2020 a 2024.	Estudios o artículos publicados fuera del período de tiempo establecido (2020 a 2024), ya que no avalarían los enfoques y tendencias actuales que aporten a la investigación.
Estudios o artículos provenientes de fuentes relevantes previamente definidas.	Fuentes de estudio que no sean académicas o científicas, como revistas no académicas, páginas web no confiables o artículos sin respaldo científico.
Estudios o artículos que presenten datos empíricos que respalden las teorías propuestas.	Estudios o artículos puramente teóricos sin un respaldo de datos que permita evaluar la efectividad de los enfoques propuestos.

3.3. Ejecución

El proceso de búsqueda se ejecutó en las bases de datos bibliográficas detalladas a continuación: Science Direct, IEEE Xplore, ACM Digital Library y Springer. Se aplicaron las búsquedas con las cadenas definidas previamente, obteniendo un total de 1 137 artículos. Estos artículos identificados fueron sometidos a los criterios de inclusión y exclusión establecidos, resultando en 448 artículos. Como siguiente filtro, se revisaron los títulos, resúmenes y palabras clave de cada estudio, obteniendo un total de 103 artículos. Finalmente, se revisaron las introducciones

y conclusiones de los artículos para determinar la relevancia de cada estudio para la investigación, seleccionando un total de 21 artículos.

En resumen, después de la revisión exhaustiva de la bibliografía (palabras clave, criterios de inclusión y exclusión definidos, resúmenes de los artículos científicos y verificación del contexto abordado en los mismos), se seleccionaron finalmente 21 artículos científicos con gran potencial para sustentar y desarrollar la investigación de esta tesis.

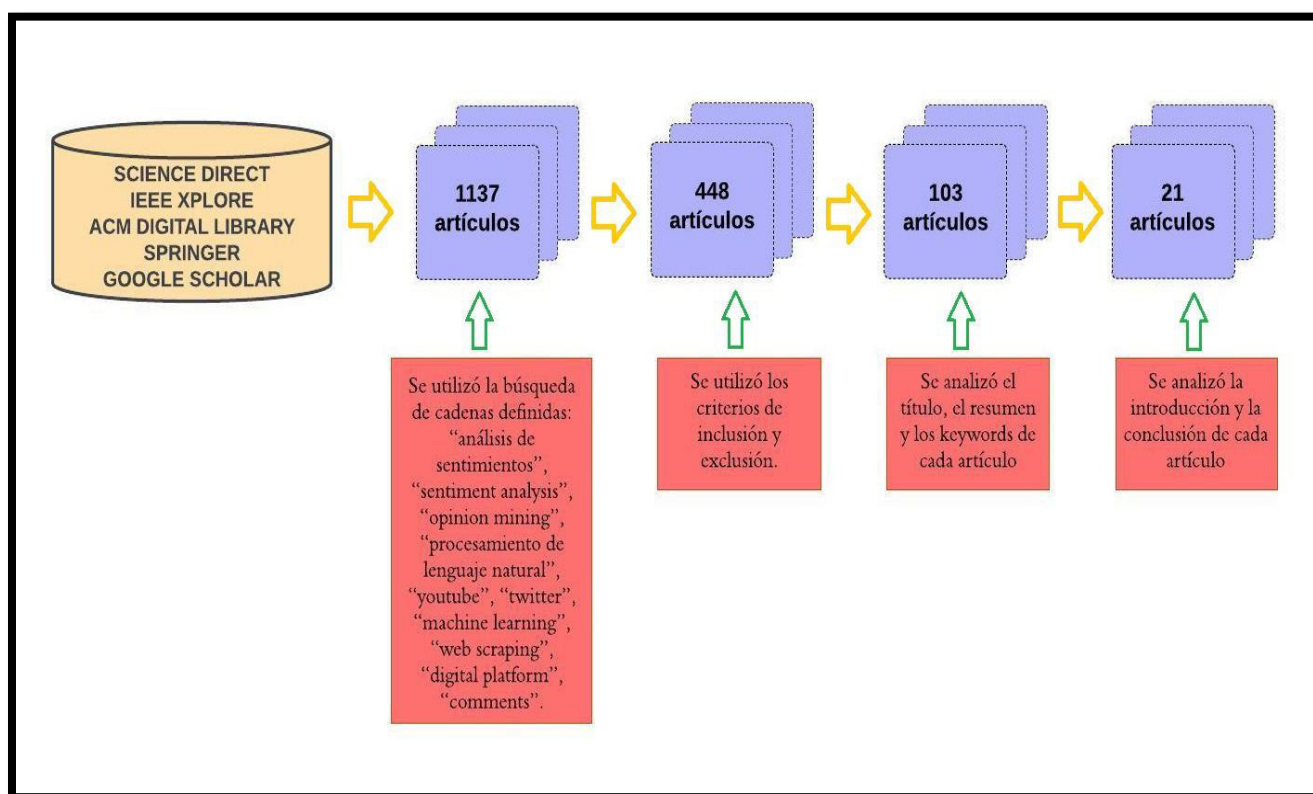


Figura 3. Proceso de revisión sistemática de la literatura para el estado del arte

Fuente: *Elaboración propia*

El proceso de selección de artículos permitió realizar un filtrado de la literatura científica más relevante, proporcionando la base teórica y metodológica necesarias para reflejar los enfoques más actuales.

3.4.Resultados

En la Tabla 3, se muestran las publicaciones seleccionadas luego del proceso de búsqueda de artículos en las diferentes bases de datos, destacadas por el aporte que ofrecerán a la investigación. Se puede observar que, de los 21 artículos científicos finalmente seleccionados, la mayoría de los journals provienen de Science Direct.

Tabla 3. Cantidad de artículos seleccionados por base de datos

Fuente: Elaboración propia

Banco de datos	Nº Journals seleccionados	Porcentaje sobre el total
Science Direct	10	47%
IEEE Xplore	6	29%
ACM Digital Library	3	14%
Springer	2	10%
TOTAL	21	100%

Como resultado de la búsqueda y el análisis realizados en los bancos de datos seleccionados (Science Direct, IEEE Xplore, ACM Digital Library y Springer), se identificaron 21 artículos científicos relevantes que aportarán significativamente a la investigación y permitirán responder las preguntas planteadas para desarrollar un aporte valioso.

A continuación, se presentan los journals seleccionados en la Tabla 4.

Tabla 4. Lista de los 21 artículos seleccionados para la investigación

Fuente: Elaboración propia

ID	Título del artículo	Autor	Journal	Año
P01	A contextual-based approach for sarcasm detection	Nivin A. Helal, Ahmed Hassan, Nagwa L. Badr, Yasmine M. Afify	Scientific Reports	2024

P02	A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM	Md Saef Ullah Miah, Md Mohsin Kabir, Talha Bin Sarwar, Mejdl Safran, Sultan Alfarhood, Mridha M. F.	Scientific Reports	2024
P03	Aspect-based drug review classification through a hybrid model with ant colony optimization using deep learning	Putta Durga, Deepthi Godavarthi, Shashi Kant, Santi Swarup Basa	Discover Computing	2024
P04	Augmenting Sentiment Prediction Capabilities for Code-Mixed Tweets with Multilingual Transformers	Ehtesham Hashmi, Sule Yildirim Yayilgan, Sarang Shaikh	Social Network Analysis and Mining	2024
P05	Cross-domain aspect-based sentiment classification with hybrid prompt	Shi Yuan, Meiqi Li, Yifei Du, Yongle Xie	Expert Systems with Applications	2024
P06	Emotion AWARE: an artificial intelligence framework for adaptable, robust, explainable, and multi-granular emotion analysis	Gihan Gamage, Daswin De Silva, Nishan Mills, Damminda Alahakoon, Milos Manic	Journal of Big Data	2024
P07	Misogynistic attitude detection in YouTube comments and replies	Aakash Singh, Deepawali Sharma, Vivek Kumar Singh	Computer Speech & Language	2024
P08	One-way ticket to the moon? An NLP-based insight on the phenomenon of small-scale neo-broker trading	Gillian Kant, Ivan Zhelyazkov, Anton Thielmann, Christoph Weisser, Michael Schlee, Christoph Ehrling, Benjamin Säfken, Thomas Kneib	Social Network Analysis and Mining	2024
P09	SNSMiner_VAC: Analyzing vaccination based on social network service data for safety surveillance	Gyun-Ho Roh, Seunghee Lee, Jong-Yeup Kim, Suehyun Lee	Expert Systems with Applications	2024
P10	StackedCNNBiLSTMCOVID: An effective stacked ensemble deep learning framework for sentiment analysis of Arabic COVID-19 tweets	Naglaa Abdelhady, Taysir Hassan A. Soliman, Mohammed F. Farghally	Journal of Cloud Computing: Advances,	2024

			Systems and Applications	
P11	Assessing Accuracy: A Study of Lexicon and Rule-Based Packages in R and Python for Sentiment Analysis	Amin Mahmoudi, Dariusz Jemielniak, Leon Ciechanowski	IEEE Access	2024
P12	Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN)	Putta Durga, Deepthi Godavarthi	IEEE Access	2023
P13	Efficient Sentiment-Aware Web Crawling Methods for Constructing Sentiment Dictionary	Byung-Won On, Jun-Young Jo, Hyunkwang Shin, Jangwon Gim, Gyu Sang Choi, Soo-Mok Jung	IEEE Access	2021
P14	Exploring E-Commerce Product Experience Based on Fusion Sentiment Analysis Method	Huaqian He, Guijun Zhou, Shuang Zhao	IEEE Access	2022
P15	Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis	Yabing Wang, Guimin Huang, Jun Li, Hui Li, Ya Zhou, Hua Jiang	IEEE Access	2021
P16	Sentiment Analysis Using Pre-Trained Language Model with No Fine-Tuning and Less Resource	Yuheng Kit y Musa Mohd Mokji	IEEE Access	2022
P17	Analyzing and Predicting Consumer Response to Short Videos in E-Commerce	Yutong Guo, Chao Ban, Jiang Yang, Khim-Yong Goh, Xiao Liu, Xixian Peng, Xiaobo Li	ACM Digital Library	2024
P18	Beyond Songs: Analyzing User Sentiment Through Music Playlists and Multimodal Data	Yipei Chen, Hua Yuen, Baojun Ma, Limin Wang, Yu Qian	ACM Digital Library	2024
P19	DOMAIN: Explainable Credibility Assessment Tools for Empowering Online Readers Coping with Misinformation	Danielle Caled, Paula Carvalho, Francisco Sousa, Mário Silva	ACM Digital Library	2024

P20	Sentiment analysis classification system using hybrid BERT models	Amira Samy Talaat	Springer	2023
P21	Sentiment analysis of tweets using text and graph multi-views learning	Loitongbam Gyanendro Singh, Sanasam Ranbir Singh	Springer	2024

Para sintetizar esta información, se elaboraron gráficas que permiten una mejor visualización de los artículos seleccionados, según el año de publicación y la base de datos especificada en los criterios de inclusión y exclusión descritos al inicio de este capítulo.

En la Figura 4, se presenta el número de estudios relacionados con el análisis de sentimiento correspondientes a los años 2021 a 2024.

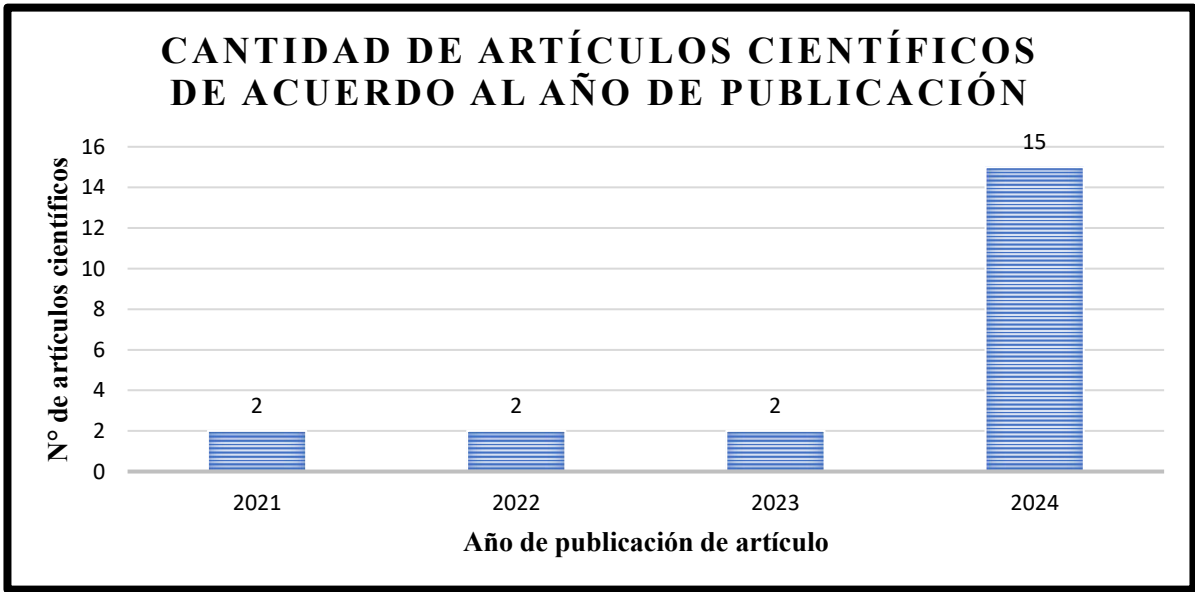


Figura 4. Gráfica con la cantidad de artículos científicos por año de publicación

Fuente: Elaboración propia

En la Figura 5, se muestran las publicaciones realizadas sobre la obtención de requisitos de las diferentes bases de datos. Es posible ver que la mayoría de los journals provienen de Science Direct, representando un 47% del total de journals tomados en cuenta.

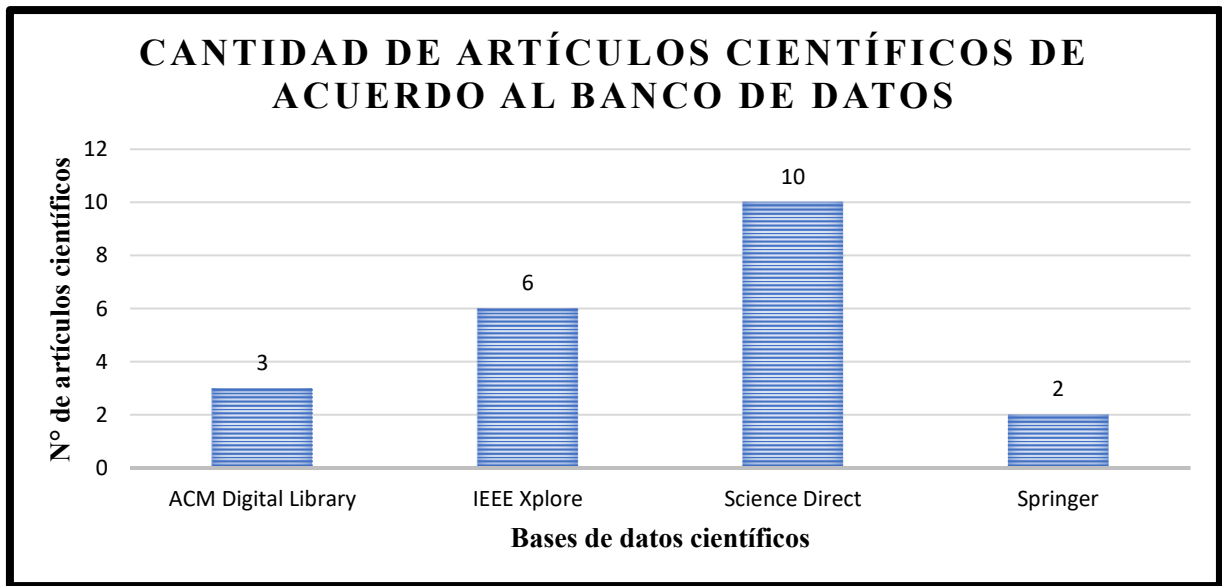


Figura 5. Gráfica con la cantidad de artículos científicos por banco de datos

Fuente: Elaboración propia

3.5. Análisis

3.5.1. ¿Qué herramientas y técnicas son las más efectivas para la recopilación y extracción de la información necesaria para el análisis de sentimientos?

Esta pregunta de investigación es fundamental, ya que definirá las herramientas y técnicas más apropiadas para obtener información fidedigna y necesaria para llevar a cabo un análisis de sentimientos efectivo en el ámbito de la investigación.

Luego de revisar los artículos seleccionados, se presenta el siguiente cuadro con las principales respuestas a esta pregunta de investigación:

Tabla 5. Herramientas y técnicas de obtención de datos para el análisis de sentimientos

Fuente: Elaboración propia

Herramienta	Descripción y/o aporte	Referencia
Datasets de Kaggle	Se utiliza un conjunto de datos de Kaggle sobre el tema requerido. Kaggle es una plataforma gratuita donde se pueden usar diversos datasets al registrarse con un usuario.	(Durga & Godavarthi, 2023) (Talaat, 2023)
Léxicos de sentimientos	Se utilizan léxicos de sentimientos, cada uno con un tamaño y rango de puntuación. Los utilizados en el artículo son: AFINN, Sentiment Lexicon, NRC Emoticon, Lexicon, SenticNet 5, SentiWordNet 3.0, etc.	(On et al., 2021) (Wang et al., 2021)
Diccionario de sentimientos	Se utilizan diccionarios de sentimientos cuando se tiene una lista de palabras, cada una con un rango de puntuación que indica su sentimiento.	(Kit & Mokji, 2022)
Dataset Mustard	“Mustard” es una dataset con una colección de escenas de comedias de televisión, con series populares como Friends y The Big Bang Theory. Este dataset combina texto, audio y video para identificar el sarcasmo.	(Helal et al., 2024)
Dataset News Headlines	“News Headlines” es un dataset que se centra en los titulares de noticias en línea, procedente de dos sitios web: ‘The Onion’ (portal que se especializa en crear versiones satíricas de la actualidad) y ‘HuffPost’ (portal que incluye titulares de noticias reales).	(Helal et al., 2024)
APIs de Redes Sociales	Una red social ofrece una API para extraer información relevante	(Durga & Godavarthi, 2023) (Chen et al., 2024)

	accediendo a esta de manera pública e integrándola a aplicaciones externas de interés.	
Web Scraping	Se utilizan herramientas de software para extraer información específica de un sitio web para su posterior uso.	(Miah et al., 2024) (Durga et al., 2024)

En conclusión, los artículos revisados destacan la utilidad de los datasets abiertos, las APIs de redes sociales y las técnicas de web scraping como métodos efectivos para la recolección de datos en contextos específicos, como YouTube, siendo estas opciones óptimas para el enfoque de la presente investigación.

3.5.2. ¿Cuáles son las técnicas más efectivas de preprocesamiento para preparar los datos en un formato adecuado para el análisis de sentimientos?

Esta pregunta de investigación es fundamental, ya que definirá los métodos más apropiados para realizar el preprocesamiento de datos, proporcionando información limpia y adecuada para su posterior análisis. Realizar este proceso garantizará la precisión de los resultados en el presente estudio.

Luego de revisar los artículos seleccionados, se presenta el siguiente cuadro con las principales respuestas a esta pregunta de investigación:

Tabla 6. Métodos de preprocesamiento de datos para el análisis de sentimientos

Fuente: Elaboración propia

Técnica	Descripción y/o aporte	Referencia
---------	------------------------	------------

Eliminación de palabras vacías	Eliminar las ‘stop words’, estas son palabras que no aportan un significado relevante. Ejemplo: y, de, la, a, etc.	(Durga & Godavarthi, 2023) (He et al., 2022) (Miah et al., 2024) (Durga et al., 2024)
Eliminación de caracteres especiales	Eliminar los símbolos, puntuaciones, números y otros caracteres que no aportan un significado relevante.	(Durga & Godavarthi, 2023) (He et al., 2022) (Miah et al., 2024) (Talaat, 2023)
Tokenización	Proceso donde se divide el texto en pequeñas unidades llamadas ‘tokens’, con el fin de facilitar el análisis del texto.	(Durga & Godavarthi, 2023) (He et al., 2022) (Kit & Mokji, 2022) (Helal et al., 2024) (Miah et al., 2024) (Durga et al., 2024)
Convertir palabras mayúsculas a minúsculas	Se requiere convertir las palabras mayúsculas a minúsculas para facilitar el análisis del texto.	(He et al., 2022)
Lematización	Proceso donde se transforma las formas conjugadas o variantes de la palabra a su forma básica.	(He et al., 2022) (Durga et al., 2024)
Derivación	Proceso donde se obtiene la raíz de la palabra eliminando prefijos y sufijos	(He et al., 2022)
Extracción de características textuales	Proceso donde se transforman los datos originales en un formato más adecuado y estandarizado para su uso y análisis posterior.	(Helal et al., 2024) (Durga et al., 2024)
Extracción de características multimedia	Proceso donde se transforma los datos originales en un formato más adecuado y estandarizado para su análisis posterior. Se aplica para audios o videos.	(Guo et al., 2024) (Chen et al., 2024)

En conclusión, la revisión evidenció que técnicas como la eliminación de ruido textual, normalización ortográfica, tokenización y lematización son ampliamente aplicadas para tareas de análisis de sentimientos, ya que una etapa previa de limpieza mejora la calidad del input y reduce la ambigüedad en los resultados. Por ello, estas opciones son consideradas para la presente investigación.

3.5.3. ¿Cuáles son las técnicas y algoritmos de clasificación de machine learning y NLP más efectivos y utilizados para el análisis de sentimientos?

Esta pregunta de investigación es fundamental, ya que definirá las técnicas y algoritmos de machine learning y NLP más utilizados y efectivos para el análisis de sentimientos. Esto se determinará evaluando el porcentaje de exactitud de cada técnica y la cantidad de datos procesados.

Luego de revisar los artículos seleccionados, se presenta el siguiente cuadro con las principales respuestas a esta pregunta de investigación:

Tabla 7. Técnicas de machine learning y NLP para el análisis de sentimientos

Fuente: Elaboración propia

Técnica o Algoritmo	Precisión	Dataset	Referencia
Análisis de Sentimiento Profundo DSA (basado en ABSA y PBSA) con una red neuronal recurrente basada en decisiones (D-RNN)	88,98 %	10000 tweets	(Durga & Godavarthi, 2023)
Support vector machine (SVM)	87,00 %	5480 reseñas	(He et al., 2022)

	62,10%	6900 reseñas	(Guo et al., 2024)
	70,99%	200 playlists	(Chen et al., 2024)
VADER	82,40 %	5480 reseñas	(He et al., 2022)
K-Means	62,56%	200 playlists	(Chen et al., 2024)
Random Forest	63,90%	6900 reseñas	(Guo et al., 2024)
	72,36%	200 playlists	(Chen et al., 2024)
KNN	86,70 %	5480 reseñas	(He et al., 2022)
	73,16%	2928 tweets	(Talaat, 2023)
CNN (Deep Learning)	66,81%	200 playlists	(Chen et al., 2024)
	77,16%	3131 tweets	(Singh & Singh, 2024)
BERT	87,30 %	5480 reseñas	(He et al., 2022)
	89,50%	7370 datos	(Helal et al., 2024)
	73,00%	No se especifica	(Hashmi et al., 2024)
	87,00%	4280 datos	(Caled et al., 2024)
	76,92%	3131 tweets	(Singh & Singh, 2024)

RoBERTa	78,92%	891 datos	(Miah et al., 2024)
	98,20%	3095 reseñas	(Durga et al., 2024)
	82,39%	2928 tweets	(Talaat, 2023)
Bibliotecas destacadas de R (sentimentr, SentimentAnalysis, syuzhet, qdap, meanr y Rsentiment)	57,00 %	No se especifica	(Mahmoudi et al., 2024)
Bibliotecas destacadas de Python (pattern, vadersentiment, flair, NLTK, TextBlob, stanza y AFFINN)	67,33 %	No se especifica	(Mahmoudi et al., 2024)

En conclusión, los modelos basados en arquitecturas profundas como DSA-D-RNN, BERT y RoBERTa han demostrado altos niveles de precisión para tareas de análisis de sentimientos, superando a técnicas clásicas como SVM o Random Forest. Por tanto, en esta investigación se optó por el uso de BETO, un modelo preentrenado en español derivado de BERT, junto con VADER y TextBlob que integran léxicos y clasificadores estadísticos previamente entrenados. Esta combinación permite una implementación efectiva y alineada con los fines de esta tesis.

3.5.4. ¿Qué métricas de validación son las más relevantes para evaluar la efectividad de los modelos de análisis de sentimientos?

Esta pregunta de investigación es fundamental, ya que definirá las mejores métricas de evaluación para los modelos de análisis de sentimientos, garantizando su calidad, precisión y validez.

Luego de revisar los artículos seleccionados, se presenta el siguiente cuadro con las principales respuestas a esta pregunta de investigación:

Tabla 8. Métricas de validación para modelos de análisis de sentimientos

Fuente: Elaboración propia

Métrica	Descripción y/o aporte	Referencia
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	(Durga & Godavarthi, 2023) (Helal et al., 2024) (Chen et al., 2024) (Caled et al., 2024) (Talaat, 2023) (Singh & Singh, 2024)
Precision	$\frac{TP}{TP + FP}$	(Mahmoudi et al., 2024) (Durga & Godavarthi, 2023) (He et al., 2022) (Helal et al., 2024) (Miah et al., 2024) (Durga et al., 2024) (Hashmi et al., 2024) (Guo et al., 2024) (Chen et al., 2024)
Recall	$\frac{TP}{TP + FN}$	(Durga & Godavarthi, 2023) (He et al., 2022) (Helal et al., 2024) (Miah et al., 2024) (Durga et al., 2024) (Hashmi et al., 2024) (Guo et al., 2024) (Chen et al., 2024)
F1 - Score	$2 * \frac{Precision_x * Recall_x}{Precision_x + Recall_x}$	(Durga & Godavarthi, 2023) (He et al., 2022) (Mahmoudi et al., 2024)

		(Helal et al., 2024) (Miah et al., 2024) (Durga et al., 2024) (Hashmi et al., 2024) (Yuan et al., 2024) (Chen et al., 2024) (Singh & Singh, 2024)
Error Rate	<i>1 - Accuracy</i>	(Durga & Godavarthi, 2023)
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2}$	(Mahmoudi et al., 2024)

En conclusión, las métricas más utilizadas en los artículos seleccionados son Accuracy, Precision, Recall y F1-Score, ya que permiten medir tanto el rendimiento de los modelos mediante aciertos y errores en la clasificación. Por esta razón, dichas métricas garantizan una evaluación objetiva y se consideran adecuadas para los fines de esta investigación.

3.5.5. ¿Qué criterios de diseño y desarrollo debe cumplir una plataforma web para garantizar que sea intuitiva, funcional y eficiente en la presentación de resultados de análisis de sentimientos a un usuario?

Esta pregunta de investigación es fundamental, ya que definirá los mejores criterios para garantizar que una plataforma web desarrollada cumpla con los estándares de usabilidad, funcionalidad y eficiencia para ofrecer resultados claros a los usuarios.

Luego de revisar los artículos seleccionados, se presenta el siguiente cuadro con las principales respuestas a esta pregunta de investigación:

Tabla 9. Criterios de desarrollo y diseño fundamentales de una plataforma web

Fuente: Elaboración propia

Criterios	Descripción y/o aporte	Referencia
Adaptabilidad en la funcionalidad de la plataforma	Los usuarios perciben la capacidad de adaptación de la plataforma, lo que facilita abordar diversas necesidades individuales.	(Caled et al., 2024) (Helal et al., 2024)
Diseño responsivo	Los usuarios pueden acceder al sistema desde ordenadores, tablets o teléfonos móviles, ya que se adapta a diferentes tamaños de pantalla.	(Gamage et al., 2024) (Caled et al., 2024)
Feedback de los usuarios	Los usuarios proporcionan comentarios que sirven como retroalimentación en el desarrollo del sistema.	(Talaat, 2023)
Interacción con los usuarios	El sistema ofrece herramientas como filtros personalizados que facilitan la navegación y comprensión de los resultados.	(Talaat, 2023) (Helal et al., 2024)
Usabilidad y diseño intuitivo	El diseño accesible facilita la interacción del usuario con la plataforma, mejorando la comprensión de los resultados.	(Talaat, 2023) (Gamage et al., 2024) (Caled et al., 2024)
Visualización de resultados	La plataforma presenta los resultados de manera clara y efectiva mediante gráficas interactivas para facilitar el análisis de mercado.	(Gamage et al., 2024) (Miah et al., 2024)

En conclusión, los artículos revisados resaltan criterios como usabilidad, diseño responsivo, visualización clara de resultados, interacción con el usuario y capacidad de adaptación funcional como aspectos clave para el diseño y desarrollo de plataformas web. Estos criterios permiten construir soluciones intuitivas y funcionales, alineadas con

los objetivos de esta investigación, al proponer una plataforma adecuada para usuarios no técnicos, es decir, personas sin mucha formación en tecnologías digitales, como ocurre en muchas PYMEs.

3.6. Conclusión sobre hallazgos en la revisión sistemática

La revisión de la literatura evidencia que son escasos los estudios orientados al desarrollo de soluciones en idioma español y aún más limitada es la atención a las necesidades de automatización en el análisis de mercado para las PYMEs en el país. Si bien existen múltiples modelos y técnicas para el análisis de sentimientos, en esta investigación se optó por emplear BETO, VADER y TextBlob debido a que son modelos preentrenados, robustos y compatibles con el idioma español, lo que permite una implementación eficiente y adaptable sin requerir entrenamiento desde cero.

Estos hallazgos respaldan la decisión de comparar el desempeño de estos tres enfoques en escenarios reales, con el fin de seleccionar e implementar el modelo más adecuado en una plataforma web orientada al análisis de comentarios en español extraídos de YouTube. Esto refuerza el valor diferencial de esta tesis como aporte académico y tecnológico, aplicable al entorno local, al proponer una solución práctica, accesible y focalizada en las necesidades de las PYMEs peruanas.

Capítulo 4: Diseño metodológico basado en CRISP-DM y SCRUM para el análisis de sentimientos

El presente capítulo tiene como finalidad describir la metodología propuesta, basada en CRISP-DM y Scrum, como marco para el diseño y desarrollo de la solución planteada: una plataforma web para el análisis de sentimientos en comentarios en idioma español de YouTube.

4.1. Descripción del método

La integración de la metodología CRISP-DM y Scrum permite abordar de manera estructurada el proceso de minería de datos y aplicar principios ágiles en el desarrollo de la plataforma web. Los modelos de proceso, como CRISP-DM y otros enfoques analíticos, son herramientas valiosas para los proyectos de ciencia de datos, pero pueden optimizarse aún más mediante la integración de enfoques ágiles. (Schröer et al., 2021)

En primer lugar, la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) fue propuesta por Daimler-Benz y NCR en el año 2000. Se define como una metodología aplicada al desarrollo de proyectos pertenecientes al campo de la minería de datos. CRISP-DM consta de un proceso estructurado de varias etapas sistemáticas e iterativas con el fin de encontrar conocimiento a partir de grandes conjuntos de datos en bruto para aportar en la toma de decisiones. (Organiche, 2020)

Como señalan Plotnikova et al. (2022), CRISP-DM consta de seis fases ejecutadas en iteraciones. La primera fase es la comprensión del negocio, que incluye la definición del problema, el alcance y la planificación. La fase 2 es la comprensión de los datos, que implica la recopilación inicial de datos, el control de calidad de los datos y la exploración de los datos. La

fase 3 es la preparación de los datos, donde se construye el conjunto de datos final a partir de los datos brutos. La fase 4 es el modelado, donde se seleccionan y aplican las técnicas de construcción de modelos. La fase 5 es la evaluación del modelo, donde se evalúan los hallazgos y se toman decisiones en función de ellos. Finalmente, la fase 6 es la fase de implementación, que implica poner en uso los modelos.

Huber et al. (2019) presentan el diagrama de CRISP-DM, donde se representa un proceso cíclico e iterativo. Las flechas indican la dependencia entre fases, permitiendo ajustes continuos, mientras que el círculo exterior simboliza la naturaleza repetitiva del proceso. Los conocimientos adquiridos al finalizar un proyecto pueden aplicarse en desarrollos futuros, permitiendo mejoras continuas y la optimización de los procesos.

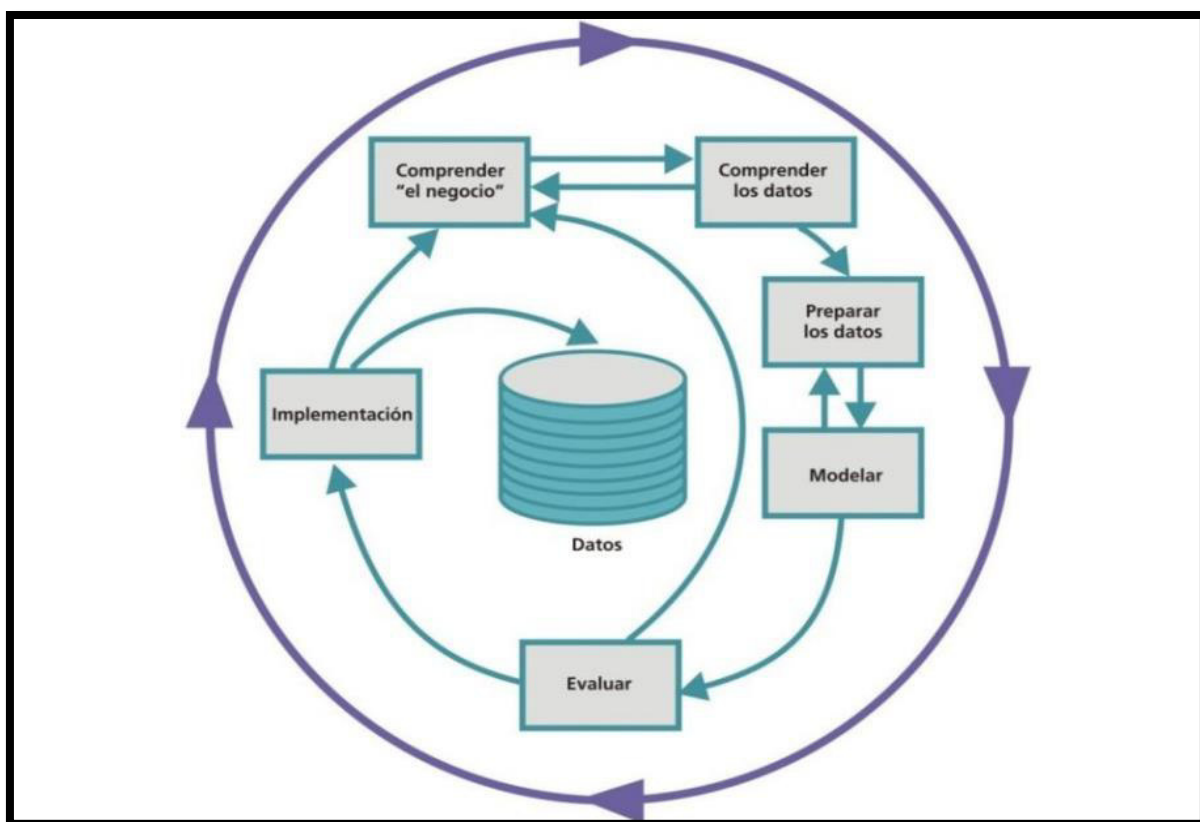


Figura 6. Procesos de la Metodología CRISP-DM

Fuente: Artículo ResearchGate (Bellini Saibene et al., 2014)

Por otro lado, tenemos el marco de trabajo Scrum, ampliamente adoptado para el desarrollo ágil de software, es un enfoque flexible, colaborativo e iterativo que permite a los equipos de desarrollo adaptarse a las preferencias cambiantes de los usuarios y a las demandas del mercado (Magnussen et al., 2024)

Tal como mencionan Mayo-Alvarez et al. (2024), Scrum se trata de un proceso iterativo que se basa en la idea de dividir un proyecto en fragmentos pequeños y manejables, llamados ‘sprints’, los cuales entregan funcionalidades, contribuyendo así al incremento del producto funcional para el cliente.

Conforme a lo expuesto por Magnussen et al. (2024), los componentes fundamentales de la metodología Scrum son:

- Roles: Product Owner, Scrum Master y Equipo de Desarrollo.
- Artefactos: Product Backlog, Sprint Backlog e Incremento.
- Eventos: Sprint Planning, Daily Scrum, Revisión y Retrospectiva del sprint.

Por tanto, en esta investigación se empleará la metodología de CRISP-DM y Scrum para implementar una plataforma web capaz de analizar los sentimientos presentes en los comentarios en español de videos de YouTube utilizando diversos enfoques de machine learning y NLP. CRISP-DM proporciona el marco base para el análisis de datos, mientras que Scrum facilita el desarrollo incremental y ágil de la plataforma web.

La Tabla 10 presenta cómo se integran estas metodologías para el desarrollo de la solución:

Tabla 10. Integración de CRISP-DM y Scrum en el desarrollo de la plataforma web

Fuente: Elaboración propia

CRISP-DM / Fase	Scrum / Sprint	Descripción
Comprensión del negocio	Sprint 0	Documento de planificación (Product Backlog)

Comprensión de los datos	Sprint 1	Extracción y el preprocesamiento de Datos
Preparación de los datos		
Modelado	Sprint 2	Desarrollo e implementación de modelos de análisis de sentimientos
Evaluación	Sprint 3	Visualización de resultados y gráficos interactivos
Despliegue	Sprint 4	Conexión del backend (API de YouTube + tratamiento de datos + modelado de análisis de sentimiento)
	Sprint 5	Diseño y desarrollo de la interfaz web (frontend)
	Sprint 6	Desarrollo del controlador (Flask)
	Sprint 7	Integración del backend, frontend y controlador (arquitectura MVC)
	Sprint 8	Pruebas de la plataforma web
	Sprint 9	Despliegue de la plataforma web
	Sprint 10	Documento final del proyecto

A continuación, se describen las etapas de CRISP-DM y Scrum consideradas como marcos de referencia para el proceso de desarrollo de la plataforma web propuesta:

4.1.1. Comprensión del negocio

En esta fase se establecen los objetivos del negocio, se comprenden los problemas existentes y se evalúan los requerimientos necesarios. Con esta información, se realiza una evaluación de la situación inicial y se procede a elaborar un plan de proyecto propuesto. (Organiche, 2020)

En la presente tesis, se identificó que muchas PYMEs peruanas presentan limitaciones para acceder a tecnologías que les permitan interpretar la percepción de sus clientes de forma automatizada. Como respuesta, se definió como objetivo de negocio el desarrollo

de una plataforma web que permita analizar sentimientos en comentarios en español de YouTube para proporcionar a estas empresas información relevante.

4.1.2. Comprensión de los datos

En esta fase se construye un conjunto de datos objetivo, seleccionando el total de datos originales o un subconjunto apropiado que sea una muestra representativa. Esta fase es muy importante porque, al ser el inicio del proceso, una adecuada selección de la muestra (que pueden ser datos, atributos o tablas) pueden mejorar el desempeño del algoritmo que se utilizará en las siguientes fases. (Organiche, 2020)

Para la presente investigación, considerando que el alcance se centra en la plataforma YouTube, se utilizó la API de YouTube Data v3 para extraer comentarios públicos de videos. Esta API permitió recolectar información estructurada como el contenido del comentario, autor, fecha de publicación, número de “me gusta” y respuestas, a través de solicitudes codificadas que cumplen con las políticas de privacidad de Google.

4.1.3. Preparación de los datos

En esta etapa se analiza la calidad de los datos obtenidos en la fase anterior para generar mejores resultados en las siguientes fases mediante la preparación y limpieza de los datos. (Organiche, 2020) Entre las tareas comunes se incluyen la eliminación de valores fuera de rango (outliers), datos nulos, información redundante y datos irrelevantes en base al objetivo planteado. Asimismo, se aplican técnicas como la reducción de dimensiones (disminución del número de variables) y transformaciones, por ejemplo, convertir valores numéricos a categóricos. (Almeida, 2024)

En el presente estudio, los comentarios en español extraídos de YouTube fueron organizados en un único conjunto de datos. Posteriormente, se aplicaron técnicas de

preprocesamiento como la eliminación de comentarios vacíos, caracteres especiales, enlaces, menciones, hashtags y palabras con números. Además, se normalizaron caracteres acentuados, se corrigieron espacios innecesarios y se unificó el texto en minúsculas. Estas acciones permitieron obtener un dataset limpio y mejorar la calidad de los datos asegurando una representación clara de la información para la siguiente fase.

4.1.4. Modelado

En esta fase del proceso se aplican técnicas o algoritmos de minería de datos para identificar patrones y extraer información relevante de los datos recopilados, alineados con los objetivos planteados en la investigación. (Organiche, 2020) Es posible que durante la ejecución de los algoritmos se deba ajustar ciertos parámetros para optimizar los resultados, los cuales varían en base al método seleccionado. En la Figura 7, se muestra como la minería de datos comprende técnicas de Machine Learning como de Procesamiento del Lenguaje Natural (NLP).

En esta investigación se implementaron tres modelos representativos de estos enfoques: VADER (modelo basado en diccionario léxico), TextBlob (modelo de tipo léxico-estadístico) y BETO (modelo de aprendizaje profundo supervisado). Finalmente, se evaluará el rendimiento de cada modelo utilizando métricas de validación, con el objetivo de determinar cuál ofrece un análisis de sentimientos más preciso y fiel al contenido de los comentarios.

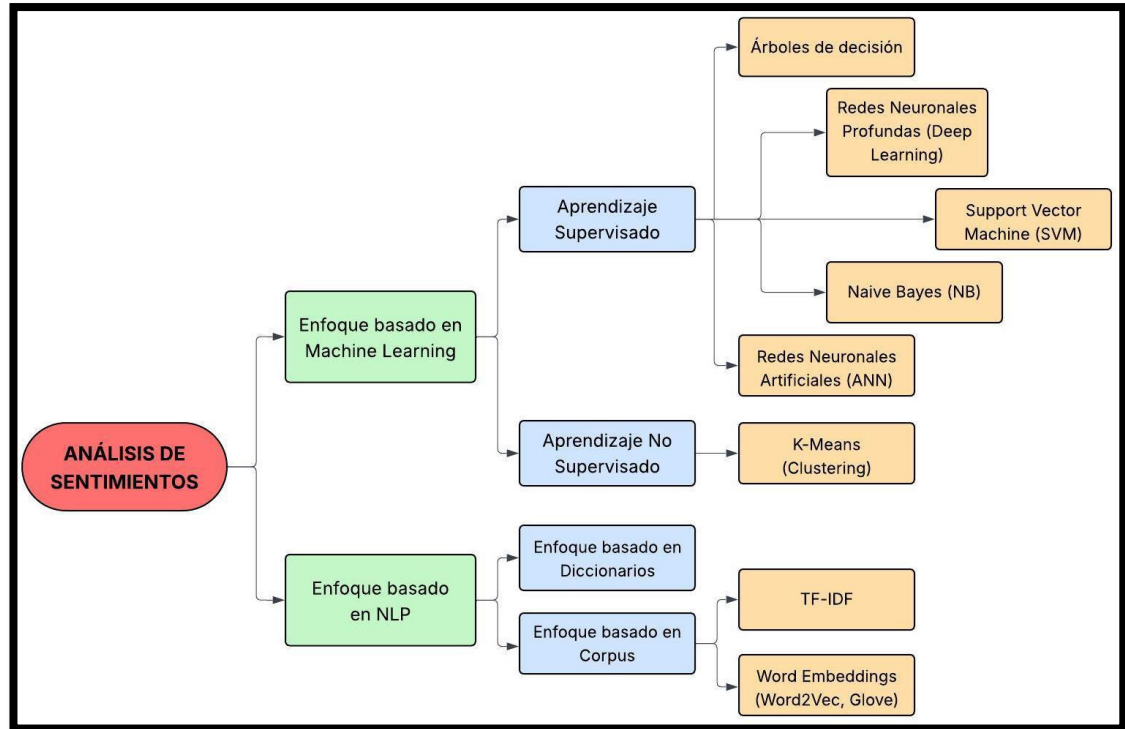


Figura 7. Técnicas y enfoques para Análisis de Sentimientos

Fuente: Elaboración propia

4.1.5. Evaluación

En esta etapa se desarrolla la interpretación y evaluación de los resultados obtenidos por los modelos aplicados, con el fin de determinar su validez, utilidad y relevancia. Para ello, se usarán métricas específicas, considerando el contexto del estudio. (Organiche, 2020)

En este estudio, se emplearán métricas de evaluación como accuracy, precision, recall y F1-score comúnmente utilizadas para medir el rendimiento de los algoritmos. Además, se aplicará la técnica de validación cruzada, que permite verificar el comportamiento del modelo ante nuevos datos evitando el sobreajuste (overfitting) y asegurando una adecuada generalización. (DataScientest, 2023)

De forma complementaria, los resultados serán representados en un formato visual comprensible que facilite su interpretación. En la tesis, la plataforma permitirá observar el porcentaje de comentarios clasificados como positivos, negativos o neutros, así como la distribución de palabras frecuentes en cada categoría mediante nubes de palabras.

4.1.6. Despliegue

Esta última fase se centra en establecer estrategias para implementar los resultados del análisis de sentimientos en un entorno intuitivo que genere valor para los usuarios. Incluye la presentación de hallazgos mediante reportes, visualizaciones y la integración de resultados en sistemas o plataformas web. Es fundamental realizar un monitoreo continuo del sistema o plataforma para garantizar su correcto funcionamiento. Considerar que en el informe final se deben redactar las experiencias y la explicación de los resultados alcanzados con la investigación. (Organiche, 2020)

En la presente investigación, los resultados del análisis de sentimientos se integrarán en una plataforma web donde el usuario podrá ingresar una URL de YouTube y visualizar los resultados mediante gráficos, nubes de palabras y reportes descargables. El despliegue de la solución considera el diseño de una interfaz responsiva, la integración del modelo seleccionado en el backend, y la generación de informes automáticos que permitan a las PYMEs interpretar fácilmente la percepción del cliente.

4.2. Selección y justificación de la metodología

En este subcapítulo se presentan las razones que respaldan la elección de las metodologías CRISP-DM y SCRUM como marcos metodológicos que guían el proceso de desarrollo de la solución propuesta, comparándolos con otras metodologías de minería de datos y desarrollo ágil.

4.2.1. Metodología para minería de datos

Contar con una metodología de minería de datos es fundamental en la investigación, ya que permite estructurar las actividades necesarias para realizar un análisis de sentimientos adecuado basado en los datos provenientes de los comentarios de un video de YouTube. Para sustentar el uso de la metodología CRISP-DM, es importante compararla con otras metodologías, tales como KDD o SEMMA. A continuación, se presenta un cuadro comparativo que contrasta la metodología CRISP-DM con las metodologías mencionadas:

Tabla 11. Comparativa de metodologías KDD, CRISP-DM y SEMMA

Fuente: Elaboración propia en base a artículo (Alaminos, A., 2023)

	KDD	CRISP-DM	SEMMA
Significado de siglas	Knowledge Discovery in Databases	Cross-Industry Standard Process for Data Mining	Sample, Explore, Modify, Model, Assess
Fases de la metodología	Consta de 7 fases: Comprensión del dominio de estudio, Selección de datos,	Consta de 6 fases: Comprensión del negocio, Comprensión de los datos,	Consta de 5 pasos: Muestreo, Exploración, Manipulación,

	Preprocesamiento de datos, Transformación de datos, Minería de Datos, Interpretación y evaluación de los resultados obtenidos y Entendimiento del conocimiento	Preparación de los datos, Modelamiento, Evaluación y Despliegue.	Modelado y Valoración.
Búsqueda de comprensión del conocimiento	Alta, asegura que el conocimiento sea útil y comprensible	Alta, asegura que el conocimiento sea útil y comprensible	Bajo, más enfoque en lo técnico
Preparación de datos	Alta, incluye creación de conjunto de datos, limpieza, preprocesamiento y transformación	Moderado, incluye comprensión de datos y preparación	Moderado, incluye muestreo, exploración y manipulación
Flexibilidad del proceso	Moderada, las fases son secuenciales	Alta, permite la iteración entre fases	Baja, tiene un flujo rígido y lineal
Implementación de nuevo conocimiento	Alta	Alta	Baja

En conclusión, CRISP-DM es la opción ideal para esta investigación debido a su estructura definida y capacidad de iteración entre fases. Su compatibilidad con Scrum permite optimizar el desarrollo de la plataforma web de análisis de sentimientos.

4.2.2. Metodología para desarrollo ágil

Contar con una metodología de desarrollo ágil es fundamental en la construcción de la plataforma web, ya que garantiza la flexibilidad y adaptabilidad del proyecto, permitiendo incorporar mejoras de manera continua durante el desarrollo. Para sustentar

el uso de la metodología Scrum, es importante compararla con otras metodologías, como Kanban. A continuación, se presenta un cuadro comparativo que contrasta la metodología Scrum con Kanban:

Tabla 12. Comparativa de metodología Scrum y Kanban

Fuente: Elaboración propia en base a artículo (Mayo-Alvarez et al., 2024)

	Scrum	Kanban
Roles	Establece los roles de: Product Owner, Scrum Master y el Equipo de Desarrollo	No tiene roles definidos
Estructura del proyecto	Uso de sprints	Tablero visual
Flexibilidad del proceso	Alta	Alta
Adaptabilidad a proyectos	Alta	Moderada,
Feedback	Retroalimentación constante	Retroalimentación constante
Planificación	Alta, se establece de manera detallada	Moderada, se establece según el avance
Riesgo de cuellos de botella	Bajo, los sprints limitan la carga de trabajo en un ciclo definido	Alto, puede detenerse el flujo debido a una acumulación de tareas si no hay control

En conclusión, Scrum es la opción ideal para esta investigación debido a su enfoque iterativo y flexible, permitiendo cambios continuos en función de las necesidades del usuario.

En definitiva, la integración de las metodologías CRISP-DM y Scrum combina el rigor estructurado del análisis de datos con la flexibilidad del desarrollo ágil, lo que la convierte en

una opción ideal tanto para el análisis de datos como para el desarrollo ágil de la plataforma web.

4.3. Esquema del desarrollo de la plataforma

Esta investigación establece una secuencia estructurada de etapas que guían la construcción del sistema de análisis de sentimientos en comentarios de YouTube como herramienta de evaluación de mercado en PYMEs.

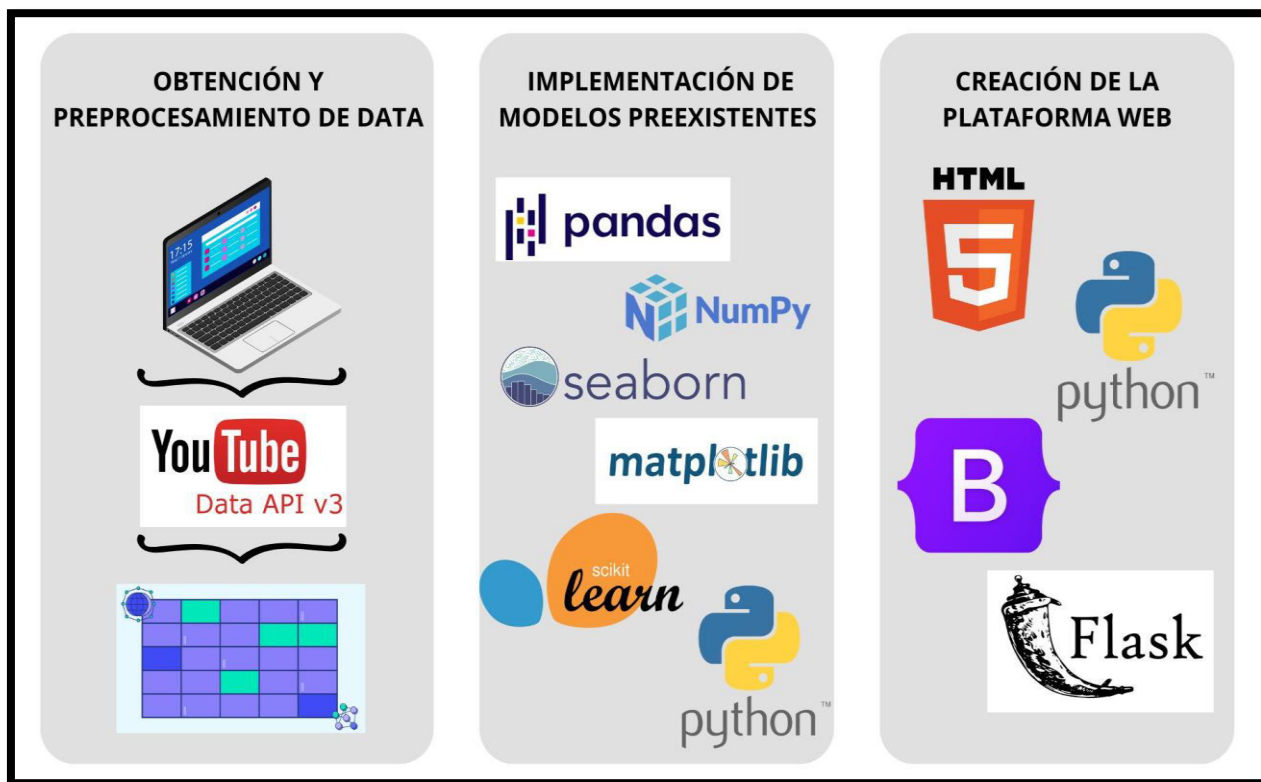


Figura 8. Esquema general del desarrollo de la plataforma web para análisis de sentimiento en comentarios de YouTube como herramienta de evaluación de mercado en PYMEs

Fuente: Elaboración propia

La Figura 8 muestra un esquema general del proceso de desarrollo de la plataforma web, el cual se encuentra alineado con los sprints definidos en la Tabla 10 según la integración metodológica planteada en este estudio.

- La fase de Obtención y Preprocesamiento de Datos (Figura 8) corresponde al Sprint 1 (Tabla 10).
- La fase de Implementación de Modelos de Análisis de Sentimientos (Figura 8) se alinea con los Sprints 2 y 3 (Tabla 10).
- La fase de Creación de la Plataforma Web (Figura 8) se desarrolla en los Sprints 4, 5, 6 y 7 (Tabla 10).
- Finalmente, las pruebas y la documentación del sistema ocurren en los Sprints 8, 9 y 10 (Tabla 10), asegurando la calidad y funcionamiento adecuado.

La estructura señalada facilita un desarrollo organizado de la plataforma web, respaldado por un marco metodológico claramente definido.

Capítulo 5: Desarrollo de la Plataforma para el Análisis de Sentimientos en comentarios de YouTube

En este capítulo se presenta de manera detallada el proceso de desarrollo e implementación del software de análisis de sentimiento propuesto, incluyendo las especificaciones, la arquitectura, las interfaces y los modelos que permitirán comprender el funcionamiento del sistema.

5.1.Requerimientos para el desarrollo del sistema

Se tomaron en cuenta los siguientes requerimientos tanto a nivel de hardware como de software para el desarrollo adecuado de la plataforma web propuesta, orientada al análisis de sentimientos en YouTube para PYMEs.

5.1.1. Hardware

Los requisitos de hardware para la implementación del sistema fueron los siguientes:

- Procesador: AMD Ryzen 7 o Intel Core i5/i7
- Capacidad de disco duro: 5 GB
- Memoria RAM: 8 GB

5.1.2. Software

Los requisitos de software para la implementación del sistema fueron los siguientes:

- Entorno de Desarrollo: Google Colab y Visual Studio Code
- Lenguajes y herramientas de desarrollo: Python 3.10 (Backend) y HTML (Frontend)
- Framework: Flask (Backend) y Bootstrap (Frontend)
- API: YouTube Data API v3

5.2.Arquitectura del sistema

Considerando los requerimientos funcionales identificados para el sistema, la plataforma cuenta con una estructura modular y escalable, lo que facilita tanto el desarrollo como el posterior mantenimiento de la aplicación. Por esta razón, la plataforma web desarrollada está basada en una arquitectura Modelo-Vista-Controlador (MVC), HTML y Bootstrap en la capa de vista, el framework Flask para la capa del controlador y Python junto con modelos especializados como VADER, TextBlob y BETO para la capa del modelo.

La elección de estas herramientas facilita la integración de nuevas funcionalidades en el futuro, considerando las necesidades cambiantes de los usuarios.

La arquitectura propuesta se muestra en la Figura 9.

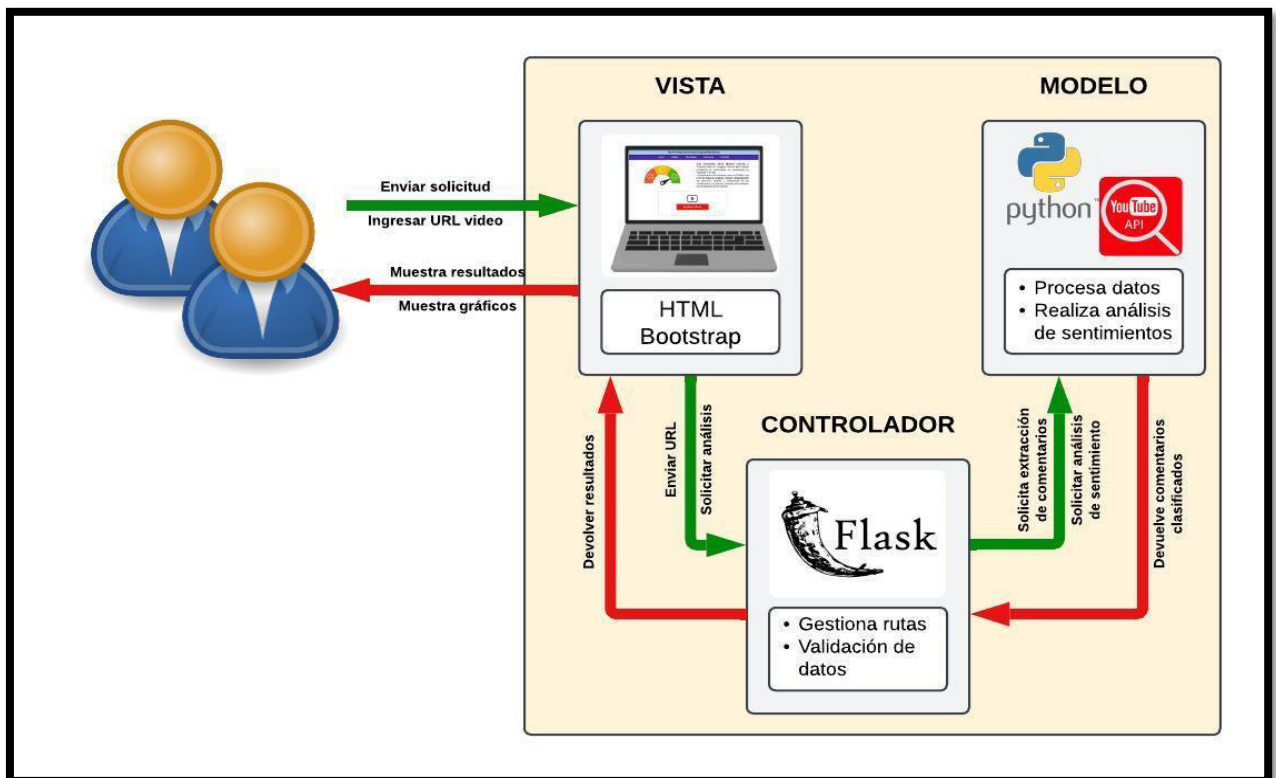


Figura 9. Arquitectura del sistema propuesto

Fuente: Elaboración propia

5.3.Descripción general del sistema

Se desarrolló un sistema diseñado para realizar un análisis de sentimientos optimizado en comentarios de YouTube en español, a través de una plataforma web intuitiva. El sistema hizo uso de librerías preentrenadas basadas en modelos de procesamiento de lenguaje natural y machine learning.

El proceso se inicia con la URL de un video de YouTube proporcionada por el usuario como dato de entrada. Como resultado, se genera una visualización detallada con información y gráficos relevantes, sintetizando los datos analizados y proporcionando un panorama claro del sentimiento predominante de los usuarios con respecto al video evaluado.

El alcance del sistema comprende lo siguiente:

- Extracción de comentarios utilizando la API de YouTube a partir de la URL de un video proporcionado por el usuario.
- Procesamiento y limpieza de los datos para garantizar la calidad de la información.
- Implementación y evaluación de tres enfoques de análisis de sentimientos (VADER, TextBlob y BETO) para garantizar la mayor efectividad y rendimiento en el análisis de sentimiento en los comentarios.
- Visualización de resultados en forma de gráfico de barras, métricas clave (número total de comentarios analizados, promedio de puntuación de sentimiento, porcentaje de sentimientos positivos / negativos / neutros, etc).
- Diseño e implementación de una interfaz accesible y responsiva que permite a los usuarios ingresar una URL de un video de YouTube y visualizar los resultados del análisis de sentimientos.

5.4.Desarrollo de la solución

5.4.1. Obtención y preprocesamiento de los datos

Para la obtención de los datos, se consideró buscar y seleccionar un dataset que contuviera los campos y variables necesarias para alcanzar el objetivo de la investigación. Sin embargo, debido a la naturaleza del análisis de sentimientos en comentarios de YouTube, no se encontró un dataset adecuado en repositorios como Kaggle, UCI Machine Learning Repository, VisualData o CMU Libraries que cumpliera con los requisitos esenciales, tales como comentarios en idioma español, datos reales de usuarios en la plataforma y alineación con el alcance de investigación para apoyar a las PYMEs.

Ante la falta de un dataset existente que cumpliera con estos criterios, se optó por la extracción directa de los datos desde videos de YouTube, utilizando la **API de YouTube Data v3**, una herramienta proporcionada por la plataforma para desarrolladores.

Este método permitió acceder a los comentarios de videos públicos de YouTube, garantizando que los datos extraídos cumplieran con los requisitos establecidos y estuvieran en conformidad con las políticas de uso de YouTube.

El proceso de obtención de datos se realizó de la siguiente manera:

- El usuario ingresa la URL del video de YouTube que desea analizar a la plataforma.
- La plataforma obtiene el ID del video mediante expresiones regulares.
- Se realiza la solicitud a la API de YouTube utilizando una clave de desarrollador válida.

- Se extraen los comentarios, respetando los límites de la API (hasta 100 comentarios por solicitud). Se verifica si hay más páginas y se continúa obteniendo comentarios y respuestas si es necesario hasta el límite permitido.
- Los comentarios extraídos se almacenan en un dataframe.

En la Tabla 13 se describen los campos obtenidos a través del API de YouTube que conformarán el dataset útil para el análisis de sentimiento en los comentarios de videos.

Tabla 13. Atributos del dataset generado a partir de la API de YouTube

Fuente: Elaboración propia

Campos Relevantes	Descripción	Tipo
Author	Nombre del usuario de YouTube que realizó el comentario en el video analizado	varchar
Comment	Contiene la opinión realizada por un usuario acerca del video analizado	varchar
Reply	Contiene la respuesta realizada por un usuario a un comentario del video analizado	varchar
Published	Fecha en la que ha sido publicado el comentario o la respuesta en el video (YYYY-MM-DD)	datetime
Likes	Cantidad de ‘Me gusta’ que ha recibido un comentario o respuesta	int
Reply Count	Número de respuestas que ha recibido un comentario	int
Reply Author	Nombre del usuario que realizó la respuesta a un comentario principal en el video analizado	varchar

El código implementado en la subfase de obtención del dataset se muestra en las Figuras 10 y 11, donde se extrae el ID del video a partir de la URL proporcionada por el usuario. Posteriormente, se ejecutó la función **get_all_youtube_comments()** para obtener los

comentarios del video, incluyendo tanto los comentarios principales como sus respuestas.

```
def extract_video_id(url):
    regex = r"(?:https?:\/\/)?(?:www\.)?(?:youtube\.com\/(?:[^\n\s]+\/\S+\/(?:v|e(?:mbded)?\/|\S*?[?&]v=[shorts\/]youtu\.be\/)([a-zA-Z0-9_-]{11}))"
    match = re.search(regex, url)
    if match:
        return match.group(1)
    else:
        return None

url = input("Por favor, ingresa la URL del video de YouTube: ")
video_id = extract_video_id(url)

if not video_id:
    print("La URL ingresada no es válida o no corresponde a un video de YouTube.")
    exit()
```

Figura 10. Código para la extracción del ID del video desde la URL

Fuente: Elaboración propia

```
def get_all_youtube_comments(video_id):
    youtube = build(YOUTUBE_API_SERVICE_NAME, YOUTUBE_API_VERSION, developerKey=DEVELOPER_KEY)
    all_comments = []

    next_page_token = None

    while True:
        response = youtube.commentThreads().list(
            part="snippet,replies",
            videoId=video_id,
            maxResults=100,
            textFormat="plainText",
            pageToken=next_page_token
        ).execute()

        for item in response["items"]:
            comment_info = item["snippet"]["topLevelComment"]["snippet"]

            comment_data = {
                "Author": comment_info["authorDisplayName"],
                "Comment": comment_info["textDisplay"],
                "Published": pd.to_datetime(comment_info["publishedAt"]).date(),
                "Likes": comment_info["likeCount"],
                "Reply Count": item["snippet"]["totalReplyCount"],
                "Reply Author": None,
                "Reply": None
            }
            all_comments.append(comment_data)

            if "replies" in item:
                for reply_item in item["replies"]["comments"]:
                    reply_info = reply_item["snippet"]
                    reply_data = {
                        "Author": None,
                        "Comment": None,
                        "Published": None,
                        "Likes": None,
                        "Reply Count": None,
                        "Reply Author": reply_info["authorDisplayName"],
                        "Reply": reply_info["textDisplay"]
                    }
                    all_comments.append(reply_data)

            next_page_token = response.get("nextPageToken")
            if not next_page_token:
                break

    return all_comments

comments = get_all_youtube_comments(video_id)
```

Figura 11. Código para la obtención de comentarios y respuestas del video de YouTube

Fuente: Elaboración propia

Para el preprocesamiento de datos, una vez obtenido el dataset con los valores en bruto (datos originales sin transformación), se procedió a preparar el dataset para mejorar la calidad de los datos de entrada y optimizar el rendimiento y la precisión de los modelos de análisis de sentimientos. En esta etapa, se identificaron y trataron datos incorrectos, incompletos o poco relevantes para la investigación.

En primer lugar, los comentarios y respuestas extraídos del dataset se integraron en un único dataframe, ya que representan exclusivamente la información clave para el análisis de sentimientos.

A continuación, se aplicaron las siguientes técnicas de preprocesamiento para procesar los comentarios y eliminar la información irrelevante:

- Eliminación de los comentarios que sean valores nulos o estén vacíos.
- Eliminación de enlaces a páginas u otros videos de YouTube.
- Eliminación de signos de puntuación y ciertos caracteres especiales.
- Eliminación de espacios innecesarios
- Eliminación de palabras que contienen números.
- Eliminación de saltos de línea.
- Eliminación de menciones a usuarios (palabras precedidas por el símbolo "@", como @usuario).
- Eliminación de hashtags (palabras o frases precedidas por el símbolo "#", como #palabra).
- Normalización de caracteres acentuados (transformación de vocales con tilde o diéresis a su forma base).
- Conversión del texto a minúsculas.

El código implementado en la subfase de preprocesamiento del dataset se muestra en la Figura 12.

```
def clean_text_round(text):
    text = text.lower()
    text = re.sub(r'<a href="https://[a-zA-Z0-9-_.]+>https://[a-zA-Z0-9-_.]+</a>', '', text)
    text = re.sub(r'\.[*?;]\%', ' ', text)
    text = re.sub(r'%s' % re.escape(string.punctuation), ' ', text)
    text = re.sub(r'\w*\d\w*', '', text)
    text = re.sub(r'<3|<33|<333', '', text)
    text = re.sub(r'\n', ' ', text)
    text = re.sub(r'@[a-zA-Z0-9-_.]+', '', text)
    text = re.sub(r'#[a-zA-Z0-9-_.]+', '', text)
    text = re.sub(r'//[a-zA-Z0-9-_.]+', '', text)
    text = re.sub(r'https://[a-zA-Z0-9-_.]+', '', text)
    text = re.sub(r'https:[a-zA-Z0-9-_.]+', '', text)
    text = re.sub(r'www.[a-zA-Z0-9-_.]+', '', text)
    text = re.sub(r'[a-zA-Z0-9-_.]+[~]', '', text)
    text = re.sub(r'[ ]+[~]', '', text)
    text = re.sub(r'\n|\\r', '', text)
    text = re.sub(r'á|ä|â|à', 'a', text)
    text = re.sub(r'é|ê|è', 'e', text)
    text = re.sub(r'í|î|ï', 'i', text)
    text = re.sub(r'ó|ô|ò', 'o', text)
    text = re.sub(r'ú|û|ù|ü', 'u', text)
    return text
```

Figura 12. Código para el preprocesamiento del dataset de comentarios de YouTube

Fuente: Elaboración propia

Con estas técnicas de preprocesamiento, se generó un dataset estandarizado y apto para la siguiente fase de análisis de sentimientos.

5.4.2. Implementación de modelos preexistentes

Para el análisis de sentimientos en comentarios de YouTube, se implementaron diferentes modelos preexistentes basados en machine learning y procesamiento de lenguaje natural (NLP). Dado el alcance de la investigación, centrado en el análisis de comentarios de videos en español, se seleccionaron modelos que ofrecen un buen procesamiento para este idioma o que pueden adaptarse desde el inglés, garantizando así una evaluación precisa del sentimiento en los comentarios analizados.

5.4.2.1. VADER

De acuerdo con la documentación oficial, VADER (Valence Aware Dictionary and sEntiment Reasoner) es un léxico y una herramienta de análisis de sentimientos basada en reglas, diseñada específicamente para interpretar los sentimientos expresados en el lenguaje informal característico de las redes sociales y otros entornos digitales. (Hutto, 2014)

Para hacer uso de VADER, se recomienda instalar la versión más reciente de nltk, ya que sus actualizaciones se gestionan a través de esta biblioteca. VADER está implementado dentro de la biblioteca NLTK en Python mediante la clase **SentimentIntensityAnalyzer**, esta clase permite obtener la polaridad de un texto sin necesidad de un entrenamiento previo. Según Hutto y Gilbert (2014), VADER ha demostrado una precisión comparable a la de evaluadores humanos ($\alpha = 0.881$) y una capacidad de clasificación superior ($F1 = 0.96$).

Es importante señalar que VADER fue diseñado para el idioma inglés, pero se decidió utilizarlo para evaluar su efectividad en el análisis de sentimientos debido a su capacidad para considerar una amplia variedad de estructuras gramaticales básicas.

El proceso de análisis de sentimientos mediante VADER se realizó de la siguiente manera:

- Se instalaron las dependencias 'nltk', 'deep_translator' y 'pandas'
- Se importaron los módulos 'deep_translator.GoogleTranslator' y 'nltk.sentiment.vader'.
- Se descargó el léxico preentrenado de VADER.

- Se definió una función para el análisis de sentimiento y se aplicó al dataframe agregando nuevas columnas de Resultado y Compound.
- VADER genera como resultado 4 valores: **neg** (puntuación de negatividad de 0 a 1), **neu** (puntuación de neutralidad de 0 a 1), **pos** (puntuación de positividad de 0 a 1) y **compound** (valor general del sentimiento de -1 a 1).

La Figura 13 muestra el código implementado de VADER para el análisis de sentimientos.

```

pip install nltk deep translator pandas
import pandas as pd
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from deep_translator import GoogleTranslator
# Descargar el léxico de VADER
nltk.download('vader_lexicon')
sid = SentimentIntensityAnalyzer()
# Función para traducir y analizar el sentimiento
def analizar_sentimiento(comentario):
    try:
        comentario_en = GoogleTranslator(source='es', target='en').translate(comentario)
        scores = sid.polarity_scores(comentario_en)
        return scores
    except:
        return {"neg": 0, "neu": 1, "pos": 0, "compound": 0} # Retorna neutral si falla

def calcular_compound(comentario):
    try:
        comentario_en = GoogleTranslator(source='es', target='en').translate(comentario)
        scores = sid.polarity_scores(comentario_en)
        return scores['compound']
    except:
        return 0 # Neutral si hay error en traducción

comentariosYTVADER['Compound'] = comentariosYTVADER['Comentario'].apply(calcular_compound)

# Filtrar el DataFrame para obtener comentarios positivos
df_positivosYTVADER = comentariosYTVADER[comentariosYTVADER['Compound'] > 0.01]
# Filtrar el DataFrame para obtener comentarios negativos
df_negativosYTVADER = comentariosYTVADER[comentariosYTVADER['Compound'] < -0.01]
# Filtrar el DataFrame para obtener comentarios neutros
df_neutrosYTVADER = comentariosYTVADER[(comentariosYTVADER['Compound'] >= -0.01) & (comentariosYTVADER['Compound'] <= 0.01)]

```

Figura 13. Código para el análisis de sentimientos VADER

Fuente: Elaboración propia

5.4.2.2. BETO

De acuerdo con la documentación oficial, BERT (Bidirectional Encoder Representations from Transformers) es un modelo de aprendizaje profundo basado en transformers. En esta investigación, se implementó una versión preentrenada llamada BETO, una variante de BERT entrenada con un corpus en español.

BETO fue desarrollado por un grupo de investigadores de la Universidad de Chile mediante el uso de la biblioteca Hugging Face Transformers. (Cañete et al., 2023)

El modelo **finiteautomata/beto-sentiment-analysis** es una versión de BETO para análisis de sentimientos, diseñada específicamente para el idioma español.

Fue entrenado con el corpus TASS 2020, que contiene aproximadamente 5000 tweets de diversos dialectos del español.

El proceso de análisis de sentimientos mediante BETO se realizó de la siguiente manera:

- Se instalaron las dependencias ‘transformers’ y ‘pandas’
- Se cargó el modelo preentrenado ‘finiteautomata/beto-sentiment-analysis’
- Se definió una función para el análisis de sentimiento y se aplicó al dataframe, agregando nuevas columnas de Resultado, Polaridad, Etiqueta y Confianza.
- BETO genera como resultado 3 valores: **POS** (Positivo), **NEU** (Neutro) y **NEG** (Negativo). Además, se tiene el valor score, que indica la confianza del modelo en su clasificación, con un valor numérico entre 0 y 1.

La Figura 14 muestra el código implementado de BETO para el análisis de sentimientos.

```

pip install transformers pandas torch
import pandas as pd
from transformers import pipeline
# Cargar el modelo entrenado para análisis de sentimientos
sentiment_analyzer = pipeline('sentiment-analysis', model="finiteautomata/beto-sentiment-analysis")
comentariosYTHUGGINGFACE=comentariosYT
def analizar_sentimiento_beto(comentario):
    try:
        resultado = sentiment_analyzer(comentario)
        label = resultado[0]['label']
        score = resultado[0]['score']

        sentimiento = {
            "POS": "Positivo",
            "NEG": "Negativo",
            "NEU": "Neutro"
        }.get(label, "Desconocido")

        return [sentimiento, label, score]

    except Exception as e:
        return ["Error", "Error", str(e)]

comentariosYTHUGGINGFACE[['Resultado', 'Etiqueta', 'Confianza']] = comentariosYTHUGGINGFACE['Comentario'].apply(
    lambda x: pd.Series(analizar_sentimiento_beto(x))
)
print(comentariosYTHUGGINGFACE[['Comentario', 'Resultado', 'Etiqueta', 'Confianza']])
comentariosYTHUGGINGFACE['Etiqueta'] = comentariosYTHUGGINGFACE['Etiqueta'].astype(str)

# Filtrar el DataFrame para obtener los comentarios según la polaridad
df_positivosYTHUGGINGFACE = comentariosYTHUGGINGFACE[comentariosYTHUGGINGFACE['Etiqueta'] == 'POS']
df_negativosYTHUGGINGFACE = comentariosYTHUGGINGFACE[comentariosYTHUGGINGFACE['Etiqueta'] == 'NEG']
df_neutrosYTHUGGINGFACE = comentariosYTHUGGINGFACE[comentariosYTHUGGINGFACE['Etiqueta'] == 'NEU']

```

Figura 14. Código para el análisis de sentimientos BETO

Fuente: Elaboración propia

5.4.2.3. TextBlob

De acuerdo con la documentación oficial, TextBlob es una biblioteca de Python que permite realizar tareas de procesamiento del lenguaje natural (NLP), tales como etiquetado de categorías gramaticales, extracción de frases nominales, análisis de sentimientos, traducción automática, etc. (Loria, 2025)

Esta biblioteca ofrece dos analizadores principales:

- NaiveBayesAnalyzer: Basado en NLTK, utiliza un clasificador de Naïve Bayes entrenado en un corpus de reseñas de películas. Para su uso, debe ser especificado explícitamente al crear un objeto. Este método requiere entrenamiento previo en base al contexto de la investigación.

- **PatternAnalyzer**: Basado en la biblioteca de Pattern, utiliza una lista de palabras con puntuaciones predefinidas de polaridad y subjetividad, lo que permite calcular el sentimiento sin requerir un entrenamiento previo. PatternAnalyzer es el analizador predeterminado en TextBlob.

Se utilizó el analizador PatternAnalyzer, ya que, aunque fue diseñado originalmente para el idioma inglés, su diccionario de palabras con polaridad predefinida ha sido adaptado a otros idiomas, permitiendo analizar los comentarios sin necesidad de traducirlos al inglés.

El proceso de análisis de sentimientos mediante TextBlob se realizó de la siguiente manera:

- Se instalaron las dependencias ‘textblob’ y ‘pandas’
- Se cargaron las bibliotecas ‘textblob’ y ‘PatternAnalyzer’
- Se define una función para el análisis de sentimiento y se aplicó al dataframe, agregando las nuevas columnas de Polaridad y Sentimiento.
- TextBlob genera como resultado un valor de polaridad, que es el valor que indicará el sentimiento, que oscilará entre -1 (muy negativo) y 1 (muy positivo).
- Luego del proceso de análisis, se establecen los siguientes parámetros: **NEGATIVO** (valor de polaridad de -1 a -0,20), **NEUTRO** (valor de polaridad -0,20 a 0,20) y **POSITIVO** (valor de polaridad de 0,20 a 1).

La Figura 15 muestra el código implementado de TextBlob para el análisis de sentimientos.

```

!pip install textblob textblob-fr pandas
import pandas as pd
from textblob import TextBlob
from textblob_fr import PatternAnalyzer

def get_polarity(text):
    if text.strip():
        tb = TextBlob(text, analyzer=PatternAnalyzer())
        polaridad = tb.sentiment[0] # Polaridad (-1 a 1)
        return round(polaridad, 3)
    return None

comentariosYTTEXTBLOB=comentariosYT
comentariosYTTEXTBLOB["Polaridad"] = comentariosYTTEXTBLOB["Comentario"].apply(lambda x: get_polarity(x) if x else None)
comentariosYTTEXTBLOB = comentariosYTTEXTBLOB.dropna(subset=["Polaridad"])
def clasificar_sentimiento(polarity):
    if polarity is None:
        return "NEUTRO"
    elif polarity <= -0.20:
        return "NEGATIVO"
    elif -0.20 < polarity <= 0.20:
        return "NEUTRO"
    else:
        return "POSITIVO"
comentariosYTTEXTBLOB["Sentimiento"] = comentariosYTTEXTBLOB["Polaridad"].apply(clasificar_sentimiento)
#Filtrar comentarios según la polaridad
df_positivosYTTEXTBLOB = comentariosYTTEXTBLOB[comentariosYTTEXTBLOB["Sentimiento"] == "POSITIVO"]
df_negativosYTTEXTBLOB = comentariosYTTEXTBLOB[comentariosYTTEXTBLOB["Sentimiento"] == "NEGATIVO"]
df_neutrosYTTEXTBLOB = comentariosYTTEXTBLOB[comentariosYTTEXTBLOB["Sentimiento"] == "NEUTRO"]

```

Figura 15. Código para el análisis de sentimientos TextBlob

Fuente: Elaboración propia

En conclusión, en esta sección se implementaron diversos modelos preexistentes para el análisis de sentimientos en comentarios en español de videos de YouTube, cada uno basado en enfoques distintos. El resultado final se determina a partir del modelo que presenta el mejor desempeño según las métricas de validación, con el objetivo de obtener una evaluación más precisa, confiable y representativa del sentimiento predominante en los comentarios analizados.

5.4.3. Creación de la Plataforma Web

Para permitir a los usuarios realizar el análisis de sentimientos en comentarios de YouTube de manera accesible y dinámica, se ha desarrollado una plataforma web que integra los modelos preentrenados mencionados anteriormente. Esta plataforma ha sido diseñada bajo la arquitectura Modelo-Vista-Controlador (MVC), tal como se definió en la sección 5.2 Arquitectura del sistema.

La plataforma cuenta con cuatro módulos específicos los cuales son:

5.4.3.1. Módulo de Extracción de Datos

En este módulo se realiza la extracción, limpieza y preparación de los comentarios obtenidos de YouTube a través del ingreso de la URL del video en un campo de texto por parte del usuario en la plataforma.

Las Figuras 16 y 17 muestran la interfaz del módulo de extracción de datos de la plataforma web.



Figura 16. Interfaz de Inicio de la plataforma web

Fuente: Elaboración propia

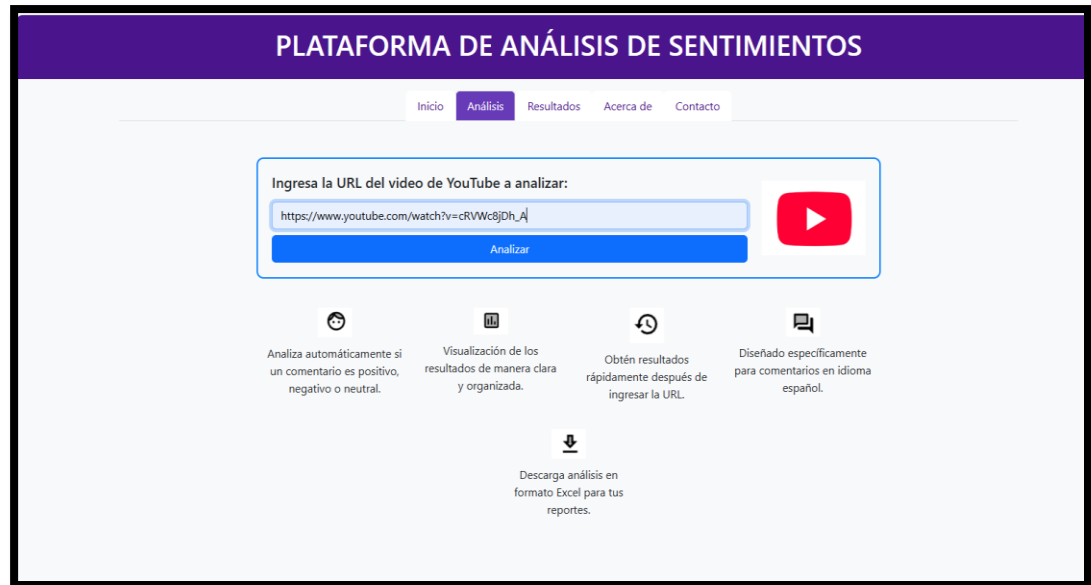


Figura 17. Interfaz de Módulo de Extracción de datos de la plataforma web

Fuente: Elaboración propia

5.4.3.2. Módulo de Análisis de Sentimientos

En este módulo se evalúa la polaridad de los comentarios del video de YouTube utilizando los modelos preentrenados VADER, BETO y TextBlob. Como se indicó previamente, cada modelo clasifica cada comentario asignándoles una etiqueta de “positivo”, “neutro” o “negativo”. Este módulo no cuenta con una interfaz visual, ya que su procesamiento ocurre en segundo plano.

La plataforma muestra los resultados basados en el análisis generado por el modelo con el mejor desempeño, identificando así el sentimiento predominante de los usuarios hacia un video de YouTube.

5.4.3.3. Módulo de Visualización de Resultados y Reportes

En este módulo se presentan los resultados del análisis de sentimientos de manera visual, interactiva y comprensible, a través de gráficos, una nube de palabras,

visualización del video analizado y la opción para descargar reportes en Excel, lo que permite una mejor interpretación de los datos.

La plataforma muestra un resumen del análisis con los siguientes indicadores:

- Cantidad total de comentarios analizados.
- Número de comentarios positivos, neutros y negativos.
- Muestra de los 10 primeros comentarios analizados del video.
- Visualización del video analizado dentro de la plataforma.

Las Figuras 18 y 19 muestran la interfaz del módulo de visualización de resultados de datos de la plataforma web.



Figura 18. Interfaz de Reportes de Resultados de la plataforma web

Fuente: Elaboración propia

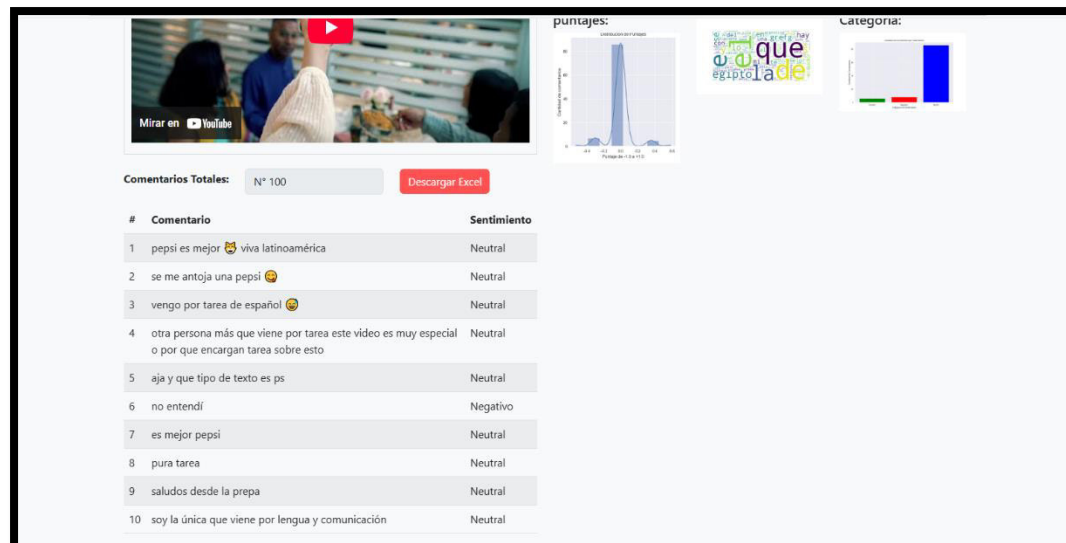


Figura 19. Interfaz de Visualización de Resultados de la plataforma web

Fuente: Elaboración propia

5.4.3.4. Módulo de Información y Contacto

Este módulo cuenta con dos secciones. La sección ‘Acerca de’ proporciona información importante sobre el objetivo de la plataforma y los autores del presente proyecto. Por otro lado, la sección de ‘Contacto’ permite a los usuarios enviar consultas o comentarios a los responsables del proyecto a través de un formulario en línea para solicitar soporte o brindar retroalimentación sobre la plataforma.

Aunque este módulo no forma parte del proceso de análisis de sentimientos, contribuye significativamente a mejorar la accesibilidad y usabilidad de la plataforma.

Las Figuras 20 y 21 muestran la interfaz del módulo de información y contacto de la plataforma web.



Figura 20. Interfaz de Información de la plataforma web

Fuente: Elaboración propia

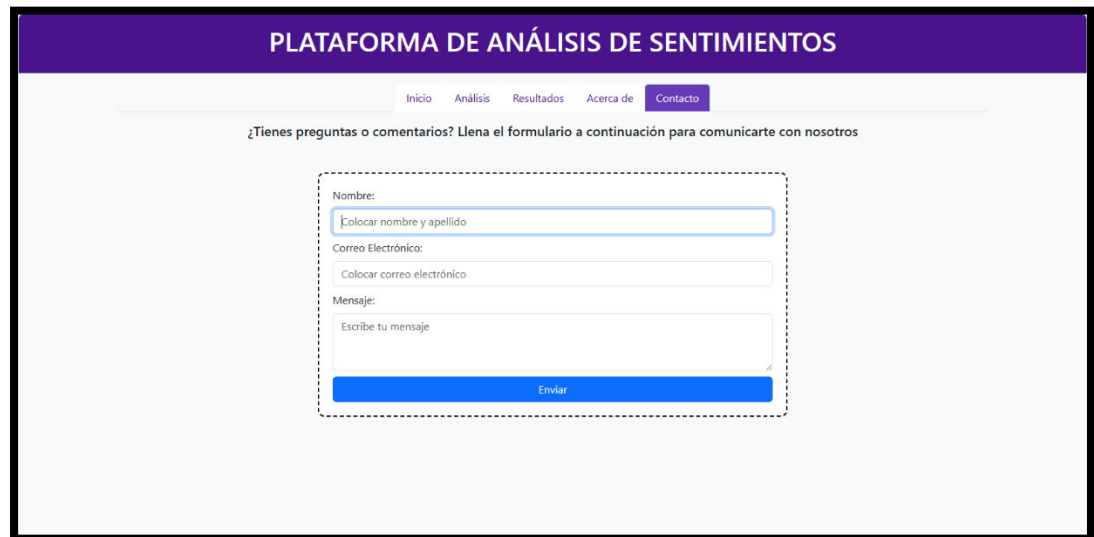


Figura 21. Interfaz de Contacto de la plataforma web

Fuente: Elaboración propia

Capítulo 6: Validación y Resultados

En este capítulo se lleva a cabo el proceso de validación de los modelos de análisis de sentimientos mediante métricas de evaluación, así como la validación de la plataforma web a través de pruebas funcionales, de usabilidad y de rendimiento. El objetivo del proceso de validación es garantizar la construcción de un producto o servicio que cumpla con las necesidades y los requisitos establecidos inicialmente por el usuario. Además, se presentan los resultados obtenidos a partir de dichas validaciones, los cuales se comparan con estudios previos y se explica el cálculo de las métricas de evaluación para evidenciar el rendimiento de los modelos en la clasificación de los comentarios.

6.1. Validación de Modelos de Análisis de Sentimientos

En esta sección se evalúa el desempeño de los modelos preentrenados de análisis de sentimientos utilizados en la plataforma web. Para ello, se aplican métricas de evaluación estándar a cada modelo con el fin de validar su funcionamiento y efectividad en el análisis de sentimientos de los comentarios de YouTube.

Para el proceso de validación de los modelos, es importante considerar los siguientes conceptos:

- Población: Conjunto total de comentarios extraídos de un video de YouTube que pueden ser analizados por la plataforma.
- Muestra: Subconjunto representativo de los comentarios extraídos de un video de YouTube. Se selecciona una muestra con el objetivo de agilizar el proceso de validación, ya que en muchos casos es inviable analizar la totalidad de los comentarios.

6.1.1. Matriz de confusión

La matriz de confusión es una herramienta representada en forma de una tabla que se utiliza para evaluar el desempeño de un modelo de clasificación. En esta investigación,

la matriz de confusión permite identificar el número de aciertos y errores en la clasificación, distribuidos en tres categorías: positivo, neutro y negativo. Dado que se manejan tres categorías, la tabla de matriz de confusión se expande a una versión de 3x3 donde las filas representan la categoría real y las columnas representan la categoría predicha por el modelo de análisis de sentimientos, esto se ilustra en la Figura 22.

		Valores Designados por Modelo		
		POSITIVO	NEUTRO	NEGATIVO
Valores Reales	POSITIVO	VERDADERO POSITIVO (TP_pos)	FALSO POSITIVO (FN_pos)	FALSO POSITIVO (FN_pos)
	NEUTRO	FALSO NEUTRO (FN_neu)	VERDADERO NEUTRO (TP_neu)	FALSO NEUTRO (FN_neu)
	NEGATIVO	FALSO NEGATIVO (FN_neg)	FALSO NEGATIVO (FN_neg)	VERDADERO NEGATIVO (TP_neg)

Figura 22. Esquema de Matriz de confusión de tres categorías

Fuente: Elaboración propia

Los resultados posibles en la matriz de confusión son los siguientes:

- Verdadero Positivo (TP_pos): El comentario es positivo y el modelo también lo clasifica como positivo.
- Falso Negativo para positivo (FN_pos): El comentario es positivo, pero el modelo lo clasifica erróneamente como neutro.

- Falso Negativo para positivo (FN_pos): El comentario es positivo, pero el modelo lo clasifica erróneamente como negativo.
- Verdadero Neutro (TP_neu): El comentario es neutro y el modelo también lo clasifica como neutro.
- Falso Negativo para neutro (FN_neu): El comentario es neutro, pero el modelo lo clasifica erróneamente como positivo.
- Falso Negativo para neutro (FN_neu): El comentario es neutro, pero el modelo lo clasifica erróneamente como negativo.
- Verdadero Negativo (TP_neg): El comentario es negativo y el modelo también lo clasifica como negativo.
- Falso Negativo para negativo (FN_neg): El comentario es negativo, pero el modelo lo clasifica erróneamente como positivo.
- Falso Negativo para negativo (FN_neg): El comentario es negativo, pero el modelo lo clasifica erróneamente como neutro.

6.1.2. Métricas de evaluación del modelo

A partir de la matriz de confusión, se pueden calcular métricas de validación para medir el rendimiento de los modelos de análisis de sentimientos propuestos. En esta investigación, para analizar los tres modelos (VADER, BETO y TextBlob), se utilizarán las métricas de accuracy, precision, recall y F1-score.

- **Accuracy:** La exactitud mide el porcentaje total de predicciones correctas en relación con el total de comentarios analizados. Para calcular la métrica **accuracy**, se utiliza la siguiente fórmula:

$$Accuracy = \frac{TP_{positivo} + TP_{neutro} + TP_{negativo}}{Total\ de\ comentarios}$$

- **Precision:** La precisión mide cuántos de los comentarios clasificados en una categoría realmente corresponden a esa categoría. Para calcular la métrica **precision**, se utiliza la siguiente fórmula:

$$Precision_x = \frac{TP_x}{TP_x + FP_x}$$

- **Recall:** El recall mide cuántos de los comentarios que realmente pertenecen a una categoría han sido correctamente clasificados en ella. Para calcular la métrica **recall**, se utiliza la siguiente fórmula:

$$Recall_x = \frac{TP_x}{TP_x + FN_x}$$

- **F1-Score:** Es la media armónica entre la métrica de precision y recall. Para calcular la métrica **F1-Score**, se utiliza la siguiente fórmula:

$$F1_x = 2 * \frac{Precision_x * Recall_x}{Precision_x + Recall_x}$$

Debemos considerar que los Falsos Positivos (FP) dependiendo de la categoría se calculan de la siguiente manera:

- **Falsos Positivos para positivo (FP_pos):** Representa la suma de valores donde la clasificación designada por el modelo sea Positivo, excluyendo los verdaderos positivos (TP_pos).

$$FP_{pos} = FN_{neu} + FN_{neg}$$

- **Falsos Positivos para neutro (FP_neu):** Representa la suma de valores donde la clasificación designada por el modelo sea Neutro, excluyendo los verdaderos positivos (TP_neu).

$$FP_neu = FN_pos + FN_neg$$

- **Falsos Positivos para negativo (FP_neg):** Representa la suma de valores donde la clasificación designada por el modelo sea Negativo, excluyendo los verdaderos positivos (TP_neg).

$$FP_neg = FN_pos + FN_neu$$

6.1.3. Proceso de validación de los modelos de análisis de sentimientos

Para evaluar el rendimiento de los modelos de análisis de sentimientos mediante las métricas mencionadas, se aplicó un proceso de validación. En primer lugar, se seleccionaron 10 videos de YouTube con diferentes temáticas, que incluyen reseñas de productos o servicios populares, campañas publicitarias de marcas y reseñas de negocios o pequeños locales. La Tabla 14 presenta los videos seleccionados para la validación, donde la columna “Total de comentarios” indica el número de comentarios extraídos mediante la API de YouTube para cada video.

Tabla 14. Lista de 10 videos utilizados para la validación del modelo

Fuente: Elaboración propia

Caso de Prueba	Categoría	Título del video	Enlace	Total de comentarios
Video #1		ELF LO BUENO Y LO MALO DE LA MARCA MAS VIRAL!	https://www.youtube.com/watch?v=FwyyHIFLGfg	530

Video #2	Reseñas de productos o servicios populares	TOP Plataformas de Streaming ACTUALIZADO 2025	https://www.youtube.com/watch?v=MwVchtpZHMmc	343
Video #3		iPhone 16 Pro-Max vs Galaxy S24 Ultra	https://www.youtube.com/watch?v=Q95MzeQDvKY	435
Video #4		Cuál es la mejor aplicación de delivery para trabajar	https://youtube.com/shorts/mMPjnbXBEQY?si=liaKnmLFZ_9hHkLo	36
Video #5	Campañas publicitarias de marcas	Yape	https://www.youtube.com/watch?v=u5_AN26BZ_g	163
Video #6		Win	https://www.youtube.com/watch?v=4K_9ftpCE28	245
Video #7		Entel	https://www.youtube.com/watch?v=V2eWRMcMiiI	177
Video #8	Reseñas de negocios o pequeños locales	FUIMOS A LA CEBICHERIA DE GASTON ACURIO: LA MAR	https://www.youtube.com/watch?v=2Qix47OA_co	837
Video #9		Así es un SUPERMERCADO en PERÚ PE	https://www.youtube.com/watch?v=ZSElclgnQDg	95
Video #10		TAMBO - Perú Una empresa que no para de crecer	https://www.youtube.com/watch?v=WsHasUWSxEo	55

Para el análisis, si el número de comentarios del video a analizar es menor o igual a 500, se analizará la totalidad de los comentarios. Sin embargo, si el número de comentarios supera los 500, se seleccionará una muestra aleatoria representativa. A continuación, se realiza un etiquetado manual de los comentarios en las tres categorías previamente definidas: positivo, neutro y negativo. Finalmente, los comentarios son procesados

utilizando los modelos VADER, TextBlob y BETO, con el fin de construir una matriz de confusión, comparar los resultados con el etiquetado y calcular las métricas de evaluación para obtener indicadores del rendimiento de cada modelo.

6.1.4. Resultado de validación de los modelos de análisis de sentimientos

Para la presentación de los resultados, estos se consolidaron en las tres categorías principales de los videos mencionados anteriormente. Se generó una matriz de confusión combinada, sumando los resultados de todos los videos dentro de cada categoría. Para cada categoría, se crearon tres matrices de confusión combinadas, correspondientes a los modelos VADER, TextBlob y BETO. Estas matrices agrupan los resultados de todos los videos dentro de cada categoría, lo que permite evaluar el desempeño de cada modelo de manera comparativa.

Se presenta en la Tabla 15, la matriz de confusión consolidada para los videos pertenecientes a la categoría de ‘Reseñas de productos o servicios populares’. En la Tabla 16, se muestra la matriz de confusión consolidada para los videos pertenecientes a la categoría de ‘Campanas publicitarias de marcas’. Finalmente, en la Tabla 17, se muestra la matriz de confusión consolidada para los videos pertenecientes a la categoría de ‘Reseñas de negocios o pequeños locales’

Tabla 15. Matriz de Confusión Consolidada para la Categoría: Reseñas de Productos o Servicios Populares

Fuente: Elaboración propia

		Valores Designados por el Modelo								
		BETO			VADER			TEXTBLOB		
		POS	NEU	NEG	POS	NEU	NEG	POS	NEU	NEG
Valores Reales	POS	291	50	4	311	25	9	17	242	85
	NEU	81	634	68	327	383	73	21	591	172
	NEG	12	41	163	82	45	89	16	128	72
	Efectividad	0.76	0.87	0.69	0.43	0.85	0.52	0.31	0.61	0.22

Tabla 16. Matriz de Confusión Consolidada para la Categoría: Campañas publicitarias de marcas

Fuente: Elaboración propia

		Valores Designados por el Modelo								
		BETO			VADER			TEXTBLOB		
		POS	NEU	NEG	POS	NEU	NEG	POS	NEU	NEG
Valores Reales	POS	169	21	24	187	20	7	8	175	31
	NEU	12	218	47	132	132	13	1	248	28
	NEG	2	11	81	21	29	44	0	68	26
	Efectividad	0.92	0.87	0.53	0.55	0.73	0.69	0.88	0.51	0.31

Tabla 17. Matriz de Confusión Consolidada para la Categoría: Reseñas de negocios o pequeños locales

Fuente: Elaboración propia

		Valores Designados por el Modelo								
		BETO			VADER			TEXTBLOB		
		POS	NEU	NEG	POS	NEU	NEG	POS	NEU	NEG
Valores Reales	POS	297	16	7	272	37	11	19	244	57
	NEU	46	392	52	209	241	40	7	396	87
	NEG	7	34	136	44	71	62	5	136	36
	Efectividad	0.85	0.89	0.70	0.52	0.70	0.55	0.61	0.51	0.20

Conclusión:

- BETO presenta un mejor desempeño global en la clasificación de comentarios, mostrando las mayores tasas de efectividad en cada categoría.
- VADER presenta una efectividad intermedia, destacándose especialmente en la clasificación de comentarios neutros.
- TextBlob presenta la efectividad más baja, enfrentando dificultad principalmente en la detección de comentarios negativos.

Posteriormente, a partir de las matrices de confusión obtenidas, se calcularon las métricas de evaluación para el análisis de sentimientos para cada modelo: exactitud, precision, recall y F1-score. Estas métricas permiten identificar cuál de los modelos ofreció el mejor rendimiento en la clasificación de comentarios según las categorías de sentimiento positivo, neutro y negativo. El cálculo se realizó considerando un conjunto

de 10 videos de prueba previamente definidos, con el fin de asegurar una evaluación objetiva y precisa de los modelos.

Tabla 18. Evaluación de métricas de validación en modelos de Análisis de Sentimientos con casos de prueba predefinidos

Fuente: Elaboración propia

	Video	Accuracy	Precision				Recall				F1-Score			
			POS	NEU	NEG	$\bar{x}$	POS	NEU	NEG	$\bar{x}$	POS	NEU	NEG	$\bar{x}$
BETO	V1	0,792453	0,79	0,83	0,65	0,75	0,83	0,82	0,60	0,75	0,81	0,82	0,62	0,75
	V2	0,801749	0,74	0,89	0,71	0,78	0,93	0,75	0,78	0,82	0,82	0,82	0,74	0,79
	V3	0,832183	0,72	0,92	0,69	0,78	0,81	0,83	0,89	0,84	0,76	0,87	0,78	0,81
	V4	0,861111	0,50	0,88	1,00	0,79	0,67	0,96	0,70	0,77	0,57	0,92	0,82	0,77
	V5	0,809815	0,55	0,88	0,77	0,73	1,00	0,79	0,82	0,86	0,71	0,83	0,80	0,78
	V6	0,751020	0,95	0,83	0,33	0,71	0,73	0,74	0,95	0,81	0,82	0,78	0,49	0,70
	V7	0,858757	0,95	0,91	0,28	0,71	0,88	0,84	1,00	0,90	0,91	0,87	0,43	0,74
	V8	0,844683	0,84	0,92	0,68	0,81	0,95	0,79	0,80	0,85	0,89	0,85	0,74	0,83
	V9	0,778947	0,89	0,76	0,77	0,81	0,67	0,86	0,71	0,74	0,76	0,81	0,74	0,77
	V10	0,800000	0,95	0,73	0,67	0,78	0,76	0,95	0,60	0,77	0,84	0,83	0,63	0,77
	$\bar{x}$	0,813072	0,79	0,86	0,66	0,77	0,82	0,83	0,79	0,81	0,79	0,84	0,68	0,77
VADER	V1	0,671698	0,56	0,86	0,53	0,65	0,91	0,60	0,36	0,62	0,69	0,71	0,43	0,61
	V2	0,577259	0,44	0,82	0,54	0,60	0,95	0,47	0,44	0,62	0,60	0,59	0,49	0,56
	V3	0,478160	0,30	0,83	0,46	0,53	0,84	0,37	0,44	0,55	0,44	0,52	0,45	0,47
	V4	0,583333	0,13	0,94	1,00	0,69	0,67	0,65	0,40	0,57	0,21	0,77	0,57	0,52
	V5	0,582822	0,12	0,72	0,84	0,56	0,83	0,60	0,54	0,66	0,20	0,65	0,65	0,50
	V6	0,620408	0,64	0,67	0,32	0,54	0,87	0,37	0,27	0,50	0,74	0,48	0,29	0,50
	V7	0,655367	0,59	0,80	1,00	0,79	0,89	0,47	0,40	0,58	0,71	0,59	0,57	0,62
	V8	0,605734	0,54	0,72	0,56	0,61	0,86	0,50	0,39	0,58	0,66	0,59	0,46	0,57

	v9	0,378947	0,18	0,53	0,63	0,45	0,67	0,47	0,15	0,43	0,29	0,50	0,24	0,34
	v10	0,581818	0,69	0,50	0,42	0,54	0,80	0,35	0,50	0,55	0,75	0,41	0,45	0,54
	$\bar{x}$	0,573554	0,42	0,74	0,63	0,60	0,83	0,49	0,39	0,57	0,53	0,59	0,46	0,52
TEXTBLOB	v1	0,507547	0,40	0,59	0,21	0,40	0,05	0,83	0,31	0,40	0,08	0,69	0,25	0,34
	v2	0,448979	0,10	0,59	0,23	0,31	0,02	0,66	0,36	0,35	0,04	0,63	0,28	0,32
	v3	0,540229	0,50	0,65	0,20	0,45	0,07	0,74	0,32	0,38	0,12	0,69	0,25	0,35
	v4	0,611111	0,50	0,74	0,36	0,53	0,33	0,74	0,40	0,49	0,40	0,74	0,38	0,51
	v5	0,619631	1,00	0,61	0,67	0,76	0,17	0,89	0,30	0,45	0,29	0,72	0,41	0,47
	v6	0,363265	0,75	0,40	0,11	0,42	0,02	0,86	0,18	0,36	0,05	0,54	0,14	0,24
	v7	0,519774	1,00	0,56	0,10	0,55	0,05	0,93	0,40	0,46	0,10	0,70	0,16	0,32
	v8	0,467144	0,71	0,52	0,20	0,47	0,05	0,82	0,23	0,37	0,10	0,64	0,21	0,32
	v9	0,442105	0,33	0,51	0,19	0,35	0,25	0,73	0,09	0,36	0,29	0,61	0,12	0,34
	v10	0,327273	1,00	0,37	0,19	0,52	0,04	0,70	0,30	0,35	0,08	0,48	0,23	0,26
	$\bar{x}$	0,484705	0,63	0,55	0,25	0,48	0,11	0,79	0,29	0,40	0,16	0,64	0,24	0,35

Conclusión:

- El modelo BETO obtuvo los mejores resultados en todas las métricas (accuracy, precision, recall y F1-score). Su exactitud promedio es 0,813, lo que indica un alto rendimiento en comparación con los otros modelos.
- El modelo VADER obtuvo un desempeño moderado con una exactitud promedio de 0,574, destacándose con una buena precisión promedio de 0,74 en la categoría neutral, aunque su desempeño en otras categorías fue deficiente.
- El modelo TextBlob obtuvo el rendimiento más bajo, con una exactitud promedio de 0,485 y F1-score de 0,35. En resumen, este no es un modelo ideal para el análisis de sentimientos de comentarios en español.

- Los modelos VADER y TextBlob muestran más variaciones de métricas entre videos lo que indica que no son consistentes en diferentes contextos.

En vista de que BETO ha demostrado ser el modelo más consistente y efectivo en todas las métricas de evaluación, se seleccionará como el modelo principal para la plataforma de análisis de sentimientos en comentarios de YouTube planteada en esta investigación.

6.2. Validación de Plataforma Web

6.2.1. Pruebas funcionales de validación

La validación funcional de la plataforma web de análisis de sentimientos es un proceso fundamental para garantizar que la plataforma cumpla con los requisitos establecidos y que cada una de sus funcionalidades se ejecute correctamente produciendo las salidas esperadas. Se verificaron diversas funcionalidades de la plataforma, tales como: ingreso de una URL de YouTube válida, la extracción de comentarios de YouTube, el análisis de sentimientos y la visualización de resultados mediante la generación de gráficas y reportes en formato Excel.

Se utilizó la herramienta Postman para evaluar estas funcionalidades a nivel del backend sin interactuar con la interfaz gráfica. Esto fue posible gracias a la arquitectura basada en Flask, que permite el manejo de solicitudes HTTP a través de endpoints internos.

Se diseñaron y ejecutaron casos de prueba que contemplan todos estos escenarios esperados, con el fin de evaluar la consistencia de la plataforma y detectar posibles fallos en la lógica de procesamiento. Con estas pruebas funcionales, se comprobó que la plataforma responde a las solicitudes del usuario y maneja adecuadamente los errores cuando se presentan datos inválidos o condiciones inesperadas.

Tabla 19. Validación funcional de los endpoints de la plataforma

Fuente: Elaboración propia

Prueba	Endpoint	Dato de Entrada	Resultado esperado	Estado
Prueba #1	/extract_comments	URL válida de video de YouTube	Lista de comentarios en formato JSON	Correcto ☑
Prueba #2	/extract_comments	URL no válida o vacía de video de YouTube	Mensaje de error “URL no válida”	Correcto ☑
Prueba #3	/analyze_sentiment	Lista de comentarios extraídos y procesados	Lista de análisis de sentimiento por comentario	Correcto ☑
Prueba #4	/generate_report	Resultados del análisis de sentimiento	Archivo descargable con reportes y gráficos	Correcto ☑

6.2.2. Pruebas de usabilidad de validación

La validación de usabilidad de la plataforma web de análisis de sentimientos es crucial para asegurar su facilidad de uso, claridad en la presentación de resultados, accesibilidad a las funciones principales y diseño responsivo adaptable a distintos dispositivos. Se realizaron las pruebas con un grupo de 15 usuarios, con edades entre los 18 y 70 años, con el objetivo de representar perfiles variados con distintos niveles de experiencia tecnológica. Los usuarios interactuaron con la plataforma y proporcionaron retroalimentación sobre su experiencia de uso.

A partir de estas pruebas de usabilidad, se identificaron áreas de mejora, como la formulación de mensajes de error y la disposición de algunos elementos visuales para mejorar la comprensión del usuario. A continuación, en la Tabla 20 se presenta una evaluación subjetiva de la usabilidad, en la cual los usuarios asignaron puntajes del 1 al

10 a diferentes criterios evaluados. Se muestra el promedio de los puntajes obtenidos para cada criterio.

Tabla 20. Resultado de evaluación subjetiva de usabilidad de la plataforma web

Fuente: Elaboración propia

Criterio Evaluado	Módulo asociado	Promedio de puntaje (1-10)	Observaciones
Ingresar la URL del video de YouTube fue una tarea sencilla y comprensible desde el inicio	Extracción de Datos	9	-
El tiempo de respuesta de la plataforma para la obtención de los comentarios fue razonable		9	Buen tiempo de obtención, aunque algunos usuarios indicaron leve demora con videos que superan los 50 mil comentarios.
El análisis de sentimiento se ejecutó correctamente y los resultados fueron coherentes	Análisis de Sentimientos	8	-
La plataforma respondió rápidamente al analizar los datos y respondió sin retrasos		7	Buen tiempo de análisis, pero con videos con muchos comentarios puede haber un ligero retraso.
Comprendí sin dificultad las categorías de análisis (positivo, neutro, negativo)		9	-
Los resultados obtenidos fueron presentados de forma clara y bien organizada	Visualización de Resultados y Reportes	8	Se sugirió agregar etiquetas en los ejes de los gráficos, lo cual fue incorporado
La funcionalidad para descargar los reportes fue fácil de identificar y utilizar		7	Se sugirió renombrar el botón de descarga a "Descargar Excel", lo cual fue implementado
La nube de palabras me permitió identificar los términos más frecuentes en los comentarios		7	-
Toda la información mostrada mantiene coherencia visual y funcional entre sus elementos		9	-
El formulario de contacto fue fácil de ubicar, completar y enviar	Información y Contacto	10	-

La plataforma fue fácil de utilizar desde el primer intento	Todos los módulos	8	Los participantes mayores de 50 años necesitaron una breve explicación inicial, se reforzó el diseño intuitivo
Pude interactuar con todas las funciones sin errores		9	-
La navegación entre secciones fue fluida e intuitiva		8	Se recomendó ubicar un menú en la parte superior para mejorar la orientación del usuario, lo cual ya fue implementado
Comprendí rápidamente el objetivo de la plataforma		8	Se incorporó la sección “Acerca de” para facilitar la comprensión del propósito de la herramienta
Las funcionalidades ofrecidas están alineadas con el objetivo principal de la herramienta		9	-
Pude visualizar correctamente la plataforma en distintos dispositivos		8	-
Confío en los resultados que genera la plataforma		8	-
El diseño gráfico de la plataforma (colores, tipografía, disposición) me pareció agradable		9	-
Recomendaría esta herramienta a otras personas o PYMEs		9	-
Considero que la plataforma es útil para PYMEs que desean conocer la percepción de sus clientes		9	-

6.2.3. Pruebas de rendimiento de validación

La validación de rendimiento de la plataforma web de análisis de sentimientos es un proceso fundamental para evaluar su eficiencia, escalabilidad y capacidad de respuesta bajo diferentes condiciones. Se realizaron pruebas con distintos videos para comprobar el desempeño de la plataforma en función de los volúmenes variables de datos. Estas pruebas de rendimiento permitieron identificar oportunidades de optimización en los

tiempos de respuesta y mejorar la capacidad para procesar grandes volúmenes de información de manera eficiente. A continuación, en la Tabla 21, se presenta la evaluación del rendimiento de los modelos BETO, VADER y TextBlob durante la extracción y el análisis de sentimientos de comentarios.

Tabla 21. Evaluación del rendimiento por modelo en extracción y análisis de comentarios

Fuente: Elaboración propia

	Video	Número de comentarios	Tiempo de extracción	Observaciones	Tiempo de análisis	Observaciones
BETO	Video #1	174	6,1450 s	Tiempo estable y proporcional al número de comentarios	22,8165 s	Buen balance entre velocidad y eficiencia del análisis.
	Video #2	1619	8,7177 s		118,1982 s	
	Video #3	6787	17,3441 s		339,0207 s	
VADER	Video #1	174	5,4092 s	Tiempo eficiente, pero muestra un aumento de tiempo con un mayor número de comentarios.	43,6629 s	Tiempo excesivo debido a la traducción automática comentario por comentario del español al inglés
	Video #2	1619	6,3950 s		362,6445 s	
	Video #3	6787	21,2863 s		1298,8206 s	
TEXTBLOB	Video #1	174	5,8949 s	El más estable y proporcional al número de comentarios.	5,1943 s	El más rápido, pero no es confiable para un análisis profundo de sentimientos debido a su baja precisión.
	Video #2	1619	7,7200 s		7,7592 s	
	Video #3	6787	16,2792 s		15,3724 s	

Capítulo 7: Conclusiones y Trabajos Futuros

7.1. Conclusión General

La presente investigación logró diseñar, desarrollar y validar una plataforma web para el análisis de sentimientos en comentarios de YouTube en español, orientada a brindar a las PYMEs una herramienta accesible que facilite la comprensión de la percepción de sus clientes sobre productos o servicios. Para su implementación se integraron los marcos metodológicos CRISP-DM y Scrum. La solución tecnológica se construyó bajo enfoques de Machine Learning y Natural Language Processing (NLP) adaptados al contexto hispanohablante, contribuyendo a reducir la brecha tecnológica que enfrentan muchas PYMEs al ofrecer una herramienta que apoya la toma de decisiones estratégicas basadas en datos confiables procesados en su idioma.

7.2. Conclusiones Específicas

- Se implementó un método eficiente para la extracción y recopilación de comentarios en español mediante la API de YouTube Data v3, con tiempos de respuesta óptimos para el posterior análisis.
- Se aplicaron técnicas de preprocesamiento de datos (eliminación de valores vacíos, ruido textual, corrección de espacios innecesarios y normalización de caracteres), incrementando la calidad del conjunto de datos.
- Se evaluaron comparativamente tres modelos de análisis de sentimientos basados en Machine Learning y NLP: VADER, TextBlob y BETO. Los resultados de la validación mostraron que BETO presentó el mejor rendimiento, alcanzando una exactitud global del

81,30% y una precisión del 86% en la clasificación de comentarios positivos, por lo que fue seleccionado como el modelo principal de la plataforma.

- Se evaluó el desempeño de los modelos mediante métricas de validación como accuracy, precision, recall y F1-score, tanto de forma global como por categoría de sentimiento.
- Se desarrolló una plataforma web con una interfaz intuitiva, multiplataforma y responsiva que presenta los resultados del análisis de sentimientos de manera clara. La plataforma superó pruebas funcionales, de usabilidad y de rendimiento, demostrando su eficacia en escenarios reales.

7.3. Trabajos Futuros

- Extender la funcionalidad de la plataforma de análisis de sentimientos para recolectar y analizar comentarios de múltiples redes sociales como Twitter, TikTok o Facebook, con el fin de obtener una visión más integral de la opinión de un producto o servicio al analizar una mayor cantidad y diversidad de comentarios de distintas plataformas.
- Incorporar la funcionalidad de análisis de sentimientos en tiempo real con actualizaciones automáticas de los resultados y alertas ante tendencias emergentes.
- Entrenar modelos de Machine Learning con un conjunto de datos más amplio y diverso en español, con el objetivo de mejorar la exactitud y adaptabilidad de los análisis a contextos locales.
- Integrar bases de datos persistentes que permitan análisis históricos y detección de cambios en la percepción de los usuarios.

- Explorar modelos multimodales que integren texto, audio y video para enriquecer el análisis de sentimientos en español. Esto permitiría, por ejemplo, reconocer emociones en la voz, interpretar el lenguaje corporal y el tono de un discurso en videos.

7.4. Recomendaciones

- Mantener actualizadas tanto las versiones del modelo BETO como las bibliotecas utilizadas (Hugging Face Transformers, NLTK) para garantizar la compatibilidad y contar con las mejoras en el rendimiento del análisis de sentimientos.
- Optimizar los tiempos de procesamiento mediante técnicas de paralelización, especialmente para realizar el análisis en videos con grandes volúmenes de datos.
- Diseñar la integración gradual de comentarios como contenidos multimedia (audio y video) de múltiples redes sociales para un análisis más integral para su aprovechamiento estratégico por parte de las PYMEs.

Referencias Bibliográficas

- Abdelhady, N., Soliman, T., & Farghally, M. (2024). Stacked-CNN-BiLSTM-COVID: an effective stacked ensemble deep learning framework for sentiment analysis of Arabic COVID-19 tweets. *Journal of Cloud Computing: Advances, Systems and Applications*, 13(1).
- Aguilar, D. & Camargo, J. (2021). *Sistema inteligente basado en redes neuronales, máquina de soporte vectorial y random forest para la predicción de deserción de clientes en microcréditos de bancos* [Tesis de título profesional, Universidad Nacional Mayor de San Marcos].
- Alaminos, A. (2023). *Introducción a la minería de texto y análisis de sentimiento con R*. Universidad de Alicante. Obets Ciencia Abierta.
- Almeida, C. (2024). *Detección de patrones académicos en curso de nivelación con deserción en la Universidad Técnica Estatal de Quevedo basado en los factores socioeconómicos* [Tesis de maestría, Universidad Técnica Estatal de Quevedo].
- Bellini, Y., Volpacchio, M., Banhero, S., & Mezher, R. (2014). Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA. *ResearchGate*.
- Caled, D., Carvalho, P., Sousa, F., & Silva, M. (2024). DOMAIN: Explainable credibility assessment tools for empowering online readers coping with misinformation. *ACM Transactions on the Web*.
- Cañete, J., Chaperon, G., Fuentes, R., Ho, J., Kang, H., & Pérez, J. (2023). Spanish pre-trained BERT model and evaluation data. *arXiv*.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000).

CRISP-DM 1.0: Step-by-step data mining guide. The CRISP-DM Consortium.

Chen, Y., Yuen, H., Ma, B., Wang, L., & Qian, Y. (2024). Beyond songs: analyzing user sentiment through music playlists and multimodal data. *ACM Transactions on Multimedia Computing Communications and Applications*.

Cooper, P. (2022). *Anuncios en YouTube: la guía completa para expertos en marketing*. Social Media Marketing & Management Dashboard. <https://blog.hootsuite.com/es/guia-publicidad-de-youtube/>

Dankar, A. & Gupta N. (2021). A systematic review of techniques, tools and applications of machine learning. *En Proceedings of the 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. IEEE.

DataScientest. (2023). *Cross-Validation: definición e importancia en Machine Learning*. Formación En Ciencia De Datos. <https://datascientest.com/es/cross-validation-definicion-e-importancia>

De Oliveira, J., Silva, A., Montalvany-Antonucci, C., Maia, G., Carvalho, K., & Macari, S. (2024). Are YouTubeTM and TikTokTM videos useful as educational tool for patients with cleft lip and palate? *Dental Press Journal of Orthodontics*, 29(6).

Durga, P., & Godavarthi, D. (2023). Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN). *IEEE Access*, 11, 108433-108447.

- Durga, P., Godavarthi, D., Kant, S., & Basa, S. (2024). Aspect-based drug review classification through a hybrid model with ant colony optimization using deep learning. *Discover Computing*, 27(1).
- Erl, T. (2008). *SOA design patterns*. Pearson Education.
- Gamage, G., De Silva, D., Mills, N., Alahakoon, D., & Manic, M. (2024). Emotion AWARE: an artificial intelligence framework for adaptable, robust, explainable, and multi-granular emotion analysis. *Journal of Big Data*, 11(1).
- Gómez, M., Vargas, J., & Vásquez, J. (2015). El estado del arte: una metodología de investigación. *Revista Colombiana de Ciencias Sociales*, 6(2), 300-320. Universidad de Antioquia.
- Guimarey, A. (2020). *Beneficios y riesgos de migrar una arquitectura monolítica a microservicios* [Tesis de maestría, Universidad de Palermo].
- Guo, Y., Ban, C., Yang, J., Goh, K., Liu, X., Peng, X., & Li, X. (2024). Analyzing and predicting consumer response to short videos in E-Commerce. *ACM Transactions on Management Information Systems*, 15(4), 1–23.
- Hashmi, E., Yayilgan, S., & Shaikh, S. (2024). Augmenting sentiment prediction capabilities for code-mixed tweets with multilingual transformers. *Social Network Analysis and Mining*, 14(1).
- He, H., Zhou, G., & Zhao, S. (2022). Exploring E-Commerce Product Experience Based on Fusion Sentiment Analysis Method. *IEEE Access*, 10, 110248-110260.
- Helal, N., Hassan, A., Badr, N., & Afify, Y. (2024). A contextual-based approach for sarcasm detection. *Scientific Reports*, 14(1).

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, 79, 403–408.

Hugging Face. (2021). *Model: finiteautomata/beto-sentiment-analysis*. Hugging Face. <https://huggingface.co/finiteautomata/beto-sentiment-analysis>

Hulla, M., Herstätter, P., Wolf, M., & Ramsauer, C. (2021). Towards digitalization in production in SMEs – A qualitative study of challenges, competencies and requirements for trainings. *Procedia CIRP*, 104, 887–892.

Hutto, C. (2014). VADER Sentiment Analysis Documentation. <https://vadersentiment.readthedocs.io/en/latest/>

Hutto, C., & Gilbert, E. (2014). *VADER: A Parsimonious rule-based model for sentiment analysis of social media text*. Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM-14), 216–225. Association for the Advancement of Artificial Intelligence (AAAI).

IBM. (2023). *¿Qué es la arquitectura orientada a servicios (SOA)?* <https://www.ibm.com/es-es/topics/soa>

Kant, G., Zhelyazkov, I., Thielmann, A., Weisser, C., Schlee, M., Ehrling, C., Säfken, B., & Kneib, T. (2024). One-way ticket to the moon? An NLP-based insight on the phenomenon of small-scale neo-broker trading. *Social Network Analysis and Mining*, 14(1).

Kit, Y., & Mokji, M. (2022). Sentiment Analysis Using Pre-Trained Language Model With No Fine-Tuning and Less Resource. *IEEE Access*, 10, 107056-107065.

- Korayem, M., Aljadda, K., & Crandall, D. (2016). Sentiment/subjectivity analysis survey for languages other than english. *arXiv*.
- Lizama, O., Kindley, G., Morales, J. & Gonzales, A. (2016). *Redes de computadores: Arquitectura cliente-servidor*. Universidad Técnica Federico Santa María.
- López, D. & Maya, E. (2017). Arquitectura de software basada en microservicios para desarrollo de aplicaciones web. *En Séptima Conferencia de Directores de Tecnología de Información, TICAL 2017: Gestión de las TICs para la Investigación y la Colaboración*. San José, Costa Rica.
- Loria, S. (2025). *TextBlob documentation (Release 0.19.0)*. <https://textblob.readthedocs.io/en/dev/>
- Magnussen, E., Havnelid, E., & Molléri, J. (2024). The product owner and its impact on success and challenges in agile Scrum projects: An Interview Study. *Procedia Computer Science*, 239, 1157–1164.
- Mahmoudi, A., Jemielniak, D., & Ciechanowski, L. (2024). Assessing accuracy: A study of lexicon and rule-based packages in R and Python for sentiment analysis. *IEEE Access*, 12, 20169-20180.
- Mayo-Alvarez, L., Del-Aguila-Arcentales, S., Alvarez-Risco, A., Sekar, M., Davies, N., & Yáñez, J. (2024). Innovation by integration of Drum-Buffer-Rope (DBR) method with Scrum-Kanban and use of Monte Carlo simulation for maximizing throughput in agile project management. *Journal of Open Innovation Technology Market and Complexity*, 10(1), 100228.

- Miah, M., Kabir, M., Sarwar, T., Safran, M., Alfarhood, S., & Mridha, M. (2024). A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. *Scientific Reports*, 14(1).
- Morales, J., Ledesma, L., & Guzmán, R. (2021). Identificación de polaridad en Twitter usando validación cruzada. *Identidad Energética*, 4.
- Nguyen, H., & Diederich, M. (2023). Facilitating knowledge construction in informal learning: A study of TikTok scientific, educational videos. *Computers & Education*, 205, 104896.
- On, B.-W., Jo, J.-Y., Shin, H., Gim, J., Choi, G., & Jung, S.-M. (2021). Efficient sentiment-aware web crawling methods for constructing sentiment dictionary. *IEEE Access*, 9, 161208-161223.
- Oracle. (2024). *What is machine learning?* Oracle. <https://www.oracle.com/pe/artificial-intelligence/machine-learning/what-is-machine-learning/>
- Organiche, E., Alfaro, A., & Barrera, G. (2020). Principales Metodologías en el Desarrollo de Proyectos de Minería de Datos. *Tecnocultura*, 51.
- Orús, A. (2024). *Número de usuarios de YouTube a nivel mundial entre 2019 y 2029*. Statista.
- Pantoja, E. (2004). El patrón de diseño modelo-vista-controlador (MVC) y su implementación en Java Swing. *Acta Nova*.
- Plotnikova, V., Dumas, M., & Milani, F. (2022). Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data & Knowledge Engineering*, 139, 102013.

Porlles, J., Yenque, J., & Lavado, A. (2006). Modelo de evaluación de mercado: Herramienta para decisiones de negocios. *Industrial Data*, 9(1), 32–39. Universidad Nacional Mayor de San Marcos.

Programa Nacional “Tu Empresa”. (2024). *Micro y pequeñas empresas representan el 99.2% del tejido empresarial peruano* [Comunicado de prensa].
<https://www.gob.pe/institucion/tuempresa/noticias/898863-micro-y-pequenas-empresas-representan-el-99-2-del-tejido-empresarial-peruano>

Python Package Index. (2024). *TextBlob: Simple Natural Language Processing*. PyPI.
<https://pypi.org/project/textblob/>

Ramaswamy, S., & DeClerck, N. (2018). Customer Perception Analysis Using Deep Learning and NLP. *Procedia Computer Science*, 140, 170–178

Reuters Institute for the Study of Journalism, University of Oxford. (2024). *Informe Digital News Report 2024*.

Richards, M. (2015). *Software architecture patterns*. O'Reilly Media.

Rodríguez, Á., Rodríguez, F., Collaguazo, D., & Rodríguez, J. (2024). Diferencias y Aplicaciones de Big Data, inteligencia artificial, machine learning y deep learning. *Dominio de las Ciencias*, 10(3).

Roh, G., Lee, S., Kim, J., & Lee, S. (2024). SNSMiner_VAC: Analyzing vaccination based on social network service data for safety surveillance. *Expert Systems With Applications*, 255, 124684.

- Schröer, C., Kruse, F., & Gómez, J. (2021). A Systematic Literature Review on Applying CRISP-DM Process model. *Procedia Computer Science*, 181, 526–534.
- Schwaber, K., & Sutherland, J. (2020). *The Scrum Guide™ – The Definitive Guide to Scrum: The Rules of the Game*. Scrum.org.
- Singh, A., Sharma, D., & Singh, V. (2024). Misogynistic Attitude Detection in YouTube comments and replies: A high-quality dataset and algorithmic models. *Computer Speech & Language*, 89, 101682.
- Singh, L., & Singh, S. (2024). Sentiment analysis of tweets using text and graph multi-views learning. *Knowledge and Information Systems*, 66(5), 2965–2985.
- Talaat, A. (2023). Sentiment analysis classification system using hybrid BERT models. *Journal of Big Data*, 10(1).
- Torrijos, C., & Goñi, A. (2022, May 25). *Del corpus al modelo de PLN*. Instituto de Ingeniería del Conocimiento (IIC).
- Wang, Y., Huang, G., Li, J., Li, H., Zhou, Y., & Jiang, H. (2021). Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis. *IEEE Access*, 9, 37075-37085.
- Yuan, S., Li, M., Du, Y., & Xie, Y. (2024). Cross-domain aspect-based sentiment classification with hybrid prompt. *Expert Systems With Applications*, 255, 124680.