



Pontificia Universidad  
Católica del Ecuador | Sede  
Ambato

## **ESCUELA DE INGENIERÍAS**

**Tema:**

**APLICATIVO SOBRE ALGORITMOS DE CLASIFICACIÓN EN RED SOCIAL**

**Proyecto de investigación previo a la obtención del título de  
Ingeniero en Sistemas de Información**

**Línea de investigación:**

**TECNOLOGÍAS DE LA INFORMACIÓN Y LA COMUNICACIÓN**

**Autor:**

Abraham Joel Mazaquiza Sánchez

**Director:**

PhD. Ricardo Patricio Medina Chicaiza

**Ambato – Ecuador**

**Agosto 2025**

## DECLARACIÓN DE AUTENTICIDAD Y RESPONSABILIDAD

Yo: **ABRAHAM JOEL MAZAQUIZA SÁNCHEZ**, con cédula de ciudadanía **1804389326**, autor del trabajo graduación titulado: "APLICATIVO SOBRE ALGORITMOS DE CLASIFICACIÓN EN RED SOCIAL", previo a la obtención del título profesional de **INGENIERO EN SISTEMAS DE INFORMACIÓN**, en la escuela de **INGENIERÍAS**.

1. Declaro tener pleno conocimiento de la obligación que tiene la Pontificia Universidad Católica del Ecuador, de conformidad con el artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de graduación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.
2. Autorizo a la Pontificia Universidad Católica del Ecuador a difundir a través del sitio web de la Biblioteca de la PUCE Ambato, el referido trabajo de graduación, respetando las políticas de propiedad intelectual de la Universidad.

Ambato, agosto 2025



Abraham Joel Mazaquiza Sánchez

CC. 1804389326

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR  
SEDE AMBATO  
APROBACIÓN DEL TRIBUNAL DE GRADO

Tema:  
APLICATIVO SOBRE ALGORITMOS DE CLASIFICACIÓN EN RED SOCIAL

Línea de investigación:

TECNOLOGÍAS DE LA INFORMACIÓN Y LA COMUNICACIÓN

Autor:

Abraham Joel Mazaquiza Sánchez

Ricardo Patricio Medina Chicaiza, Ing. PhD.  
CC. 1802333276

**CALIFICADOR**

f. 


Enrique Xavier Garcés Freire, Ing. Mg.

**CALIFICADOR**

f. 

Liliana del Rocío Mena Hernández, Ing. Mg.

**CALIFICADOR**

f. 

Darío Javier Robayo Jácome, Ing. Mg.  
**DIRECTOR ESCUELA DE INGENIERÍAS**

f. 

Diego Gonzalo Coca Chanalata, Dr.  
**SECRETARIO GENERAL PUCESA**

f. 

Pontificia Universidad  
Católica del Ecuador  
SECRETARIA GENERAL  
PROCURADURIA

Ambato – Ecuador  
Agosto 2025

## DEDICATORIA

Quiero dedicar este trabajo a mi padre Carlos Mazaquiza y a mi madre Verónica Sánchez quienes, con su amor incondicional, sus sacrificios me dieron la fuerza para no rendirme jamás. Cada consejo y cada sonrisa de orgullo fueron el combustible que me impulso hasta esta meta.

A mi hermana Karla Mazaquiza por ser mi apoyo incondicional, mi crítica más honesta y mi risa en los momentos más difíciles; sin tu complicidad y tu cariño este logro no sería el mismo. Este triunfo es tanto suyo como mío, porque ustedes son mi razón, mi inspiración y mi mayor orgullo.

## **AGRADECIMIENTO**

A Dios, por darme la vida, fuerza y sabiduría para alcanzar esta meta; a la universidad y a mis maestros, por brindarme conocimientos valiosos, oportunidades de crecimiento especialmente al área de Tecnologías de información donde realice mis prácticas preprofesionales, las cuales enriquecieron significativamente mi desarrollo profesional; y de manera muy especial a mi tutor de tesis PhD. Ricardo Patricio Medina Chicaiza cuya orientación y paciencia fueron fundamentales para la culminación exitosa de este trabajo.

Por último, a mis amigos y compañeros quienes, con su compañía sincera, sus palabras de aliento en los momentos cruciales y esas risas espontáneas que aliviaban el estrés transformaron este desafío académico en una experiencia memorable.

## RESUMEN

Esta investigación abordó el problema de los comentarios dirigidos hacia la Pontificia Universidad Católica del Ecuador Sede Ambato (PUCESA) en la red social Facebook, donde no existían herramientas automáticas que permitieran identificar este tipo de contenido. Ante esta situación se planteó como objetivo general la construcción de un aplicativo sobre algoritmos de clasificación en red social, que facilitara la detección de textos agresivos al priorizar los casos más urgentes mediante alertas.

El proyecto se desarrolló utilizando técnicas de minería de texto con el lenguaje de programación R aplicando librerías especializadas como tm, tidytext y syuzhet. Se procesaron comentarios extraídos de Facebook mediante técnicas como tokenización, eliminación de palabras vacías y transformación de texto. Posteriormente se implementaron algoritmos de clasificación como Naive Bayes y SVM para identificar el nivel de agresividad presente, los resultados se integraron en un entorno visual desarrollado con Shiny que incluyó gráfico de barras, nubes de palabras, análisis de emociones y listados de términos ofensivos más frecuentes.

El aplicativo no solo permitió identificar patrones de lenguaje agresivo, sino que también constituyó una herramienta útil para monitorear el clima emocional de la comunidad universitaria. Su uso permitió a la universidad contar con una base de análisis objetivo para anticipar posibles crisis de comunicación y tomar decisiones informadas. La validación técnica se llevó a cabo con el personal del área de marketing de la PUCESA, quienes calificaron el aplicativo obteniendo un 99.56% de calidad.

**Palabras clave:** minería de texto, análisis de emociones, clasificación de textos, lenguaje r.

## ABSTRACT

*This research addressed the problem of comments directed at the Pontificia Universidad Católica del Ecuador Sede Ambato (PUCESA) on the social network Facebook, where there were no automatic tools to identify this type of content. Given this situation, the overall objective was to build an application based on social network classification algorithms that would facilitate the detection of aggressive texts by prioritizing the most urgent cases through alerts.*

*The project was developed using text mining techniques with the R programming language, applying specialized libraries such as tm, tidytext, and syuzhet. Comments extracted from Facebook were processed using techniques such as tokenization, stop word removal, and text transformation. Classification algorithms such as Naive Bayes and SVM were then implemented to identify the level of aggression present. The results were integrated into a visual environment developed with Shiny that included bar charts, word clouds, emotion analysis, and lists of the most frequent offensive terms.*

*The application not only identified patterns of aggressive language, but it also served as a useful tool for monitoring the emotional climate of the university community. Its use provided the university an objective basis for analysis to anticipate potential communication crises and make informed decisions. Technical validation was carried out with PUCESA marketing staff, who rated the application at 99.56% quality.*

**Keywords:** *text mining, sentiment analysis, text classification, r language.*

## ÍNDICE GENERAL DE CONTENIDOS

DECLARACIÓN DE AUTENTICIDAD Y RESPONSABILIDAD .....	ii
APROBACIÓN DEL TRIBUNAL DE GRADO.....	iii
DEDICATORIA.....	iv
AGRADECIMIENTO.....	v
RESUMEN .....	vi
ABSTRACT .....	vii
INTRODUCCIÓN .....	1
CAPÍTULO I. ESTADO DEL ARTE Y LA PRÁCTICA.....	7
1.1. Principales aportes del sistema operativo: Surgimiento y Desarrollo .....	7
1.2. Comparativa de sistemas operativos para procesamiento de datos textuales	9
1.3. Minería de texto como parte de la inteligencia artificial .....	9
CAPÍTULO II. DISEÑO METODOLÓGICO .....	15
2.1. Caracterización de institución.....	15
2.2. Metodología de investigación .....	17
2.3. Fases del desarrollo del aplicativo .....	20
CAPÍTULO III. IMPLEMENTACIÓN Y VALIDACIÓN DEL APLICATIVO DE CLASIFICACIÓN DE TEXTOS AGRESIVOS PARA REDES SOCIALES.....	33
3.1. Pruebas de funcionalidad de la aplicación .....	33
3.2. Validación de la aplicación .....	38
CONCLUSIONES.....	40
RECOMENDACIONES .....	41
BIBLIOGRAFÍA .....	42
ANEXOS .....	45

## INTRODUCCIÓN

Como señala Gil Pascual (2021) La minería de texto se define como un procedimiento para obtener información conveniente a partir de datos textuales no estructurados a través de métodos computacionales y estadísticos. Esta disciplina posibilita convertir grandes cantidades de texto en conocimiento perceptible, detectando tendencias, patrones y vínculos desconocidos en los datos. Su beneficio abarca diversas áreas, desde la clasificación de documentos hasta el análisis de opiniones, siendo: una herramienta clave en la investigación y el estudio de datos textuales.

Los siguientes trabajos sobresalen en el **contexto internacional**:

(Kassambara, 2017) Afirma que el análisis de conglomerado es una técnica vital en el aprendizaje no supervisado, destacando su capacidad para reconocer patrones y agrupar elementos similares en grandes conjuntos de datos, sin requerir ningún método previo. El científico demuestra que el software R y las herramientas como FactoExtra facilitan estas tareas al proporcionar características avanzadas para ejecutar algoritmos, mostrar resultados y confirmar los resultados. Esta técnica es muy útil para académicos y expertos en diferentes campos. Se cree que el método es útil para implementar métodos complejos, pero podría mejorarse con una discusión más extensa sobre cómo manejar información fuerte o subjetiva.

(Rubio, 2019) destaca que las estadísticas de aprendizaje no deben limitarse al uso de herramientas y métodos, sino que deben incluir una comprensión exhaustiva de las ideas y su uso práctico en situaciones, una tarea en la que muchos alumnos enfrentan desafíos que este libro busca para facilitar el aprendizaje usando el software R, una herramienta sin costo y flexible que ayuda a hacer trabajo estadístico, hacer gráficos y examinar las conexiones de datos. Se convierte en una opción sólida en comparación con el software de negocios, porque es fácil de usar, flexible y respaldado por un gran grupo de académicos. Por esta razón, el libro combina la teoría estadística con ejemplos de la vida real en R, para alentar el uso de simulación y visualización gráfica de datos como métodos cruciales para

comprender los conceptos fundamentales de las estadísticas. Por esta razón, este método es útil, proporciona acceso a herramientas avanzadas, a pesar de que podría ser más ventajoso si se concentra más en las limitaciones de software y la creación de usuarios para evitar su uso básico.

(Fernández, 2023), tanto el análisis de minería de texto como el análisis de sentimientos son métodos esenciales para obtener patrones y emociones de información textual producida en redes sociales, blogs y otros medios digitales, este proyecto enfrenta el desafío de manejar grandes cantidades de texto sin estructura para ofrecer información valiosa y optimizar la toma de decisiones, los métodos de minería de texto, como la limpieza, la calculación (la calculación (la calificación de la física de reemplazo, por el procedimiento, el procedimiento de la recaudación de la identificación de la identificación que tiene la información de la decisión, la realización de la creación de la vegetación, la limpieza, la limpieza, la medición de la medición de la vida de la reclusión (el procedimiento de reemplazo por exclusivos. de los datos o recursos sin poner en riesgo su protección). Y el análisis de los sentimientos, hacer posible convertir información textual en conocimiento útil, simplificando su uso en diferentes áreas. Además, el lenguaje R ofrece una amplia variedad de herramientas y colecciones que facilitan el uso de técnicas de análisis de texto y evalúan las emociones. Herramientas como TM, TidyText y Sentiment ayudan en el funcionamiento suave del reprocesamiento de datos de texto, el examen y la representación gráfica. Estas habilidades hacen de una selección común entre los analistas de ciencias sociales e investigadores de datos. Por esta razón, este método es crucial para tareas como examinar mensajes hostiles en plataformas sociales, ofrece herramientas prácticas para su uso, pero podría ampliar la discusión sobre cuestiones morales conectadas.

Los siguientes trabajos sobresalen en el **contexto nacional**:

(Moreno Guzmán, 2021), demuestra que la evidencia del análisis de las redes sociales y el análisis de sentimientos son herramientas importantes para conocer las posiciones y comportamientos de los usuarios en sitios web como *Twitter*, que de otro modo no estaban organizados, así como métodos estadísticos El escritor

aborda el desafío de hacer frente a la serie de informes que se publican en los sitios web de las redes sociales diariamente, entre temas como el ruido en los datos y las ambigüedades idiomáticas que impiden la detección de patrones significativos. Se deduce que las técnicas de purificación, estandarización y categorización en la gestión de datos, así como R, ayudan a la identificación de emociones y sentimientos en los clientes, facilitando la toma de decisiones en campos como la política y la empresa. El escritor proporciona una solución concreta a los problemas comunes de minería de texto. Sin embargo, estoy de acuerdo en que un enfoque más sistemático para el análisis de datos, como el sarcasmo, puede mejorar los resultados logrados.

(Morán López, 2022), estuvo involucrado en su estudio sobre la comparación de plataformas para el análisis de emociones, y se investigaron aplicaciones como *Monkeyarn*, *Google Collaboratory* y *RapidMiner* para la administración de grandes conjuntos de datos. Se llevó a cabo un análisis para *Google Collaboratory* utilizando *Python* y bases de datos como *Pandas*, *WordCloud* y *Matplotlib* con la intención de extraer, categorizar y mostrar datos textuales. El autor descubrió que la elección de plataformas y algoritmos influyó en la calidad de la sensación, mostrando las variaciones en los resultados y la precisión de las herramientas comparativas. *Monkeylearn* pudo proporcionar resultados más consistentes y confiables que el etiquetado de datos manuales, gracias a su expansión en los procesos y la alta usabilidad. *Google Collaboratory* fue único porque podría funcionar con las bibliotecas avanzadas de Python. Posible: según mi opinión, esta investigación se beneficia de las muchas características y funciones de Python, que también son muy flexibles. Aunque no usaré esta palabra en mi tesis, todavía la usaré porque encaja con mis creencias académicas y tendré la oportunidad de usar herramientas especializadas. La minería de texto puede ayudarnos a comprender mejor las emociones analizando el texto escrito, y estos libros ofrecen métodos útiles para crear representaciones visuales y realizar análisis estadísticos. Además, su ubicuidad en la investigación académica asegura los hallazgos que cumplen con los estándares científicos.

(Espinosa, 2023), llevó a cabo una investigación de investigación para evaluar el impacto de la tecnología de *chatbot* en los académicos, empleando a Python debido a su adaptabilidad y la disponibilidad de bibliotecas especializadas como pandas, los escritores notaron que las conversaciones con el *chatbot* dieron como resultado una gran cantidad de datos escritos, que requerían buenos métodos para predecir, clasificar y mostrar los estados de ánimo y los sentimientos. Se descubrió que el uso de Python y sus librerías especializadas ayudaron a administrar bien los datos, clasificando con precisión las respuestas de los participantes de la encuesta y ayudando a detectar tendencias emocionales importantes. Se reconoce como un estudio en el que el enfoque adoptado se reconoce positivamente, Python es comúnmente conocida por su adaptabilidad y eficiencia en el análisis de datos. A pesar de esto, optaré por el lenguaje R en mi proyecto, se reconoce en la comunidad académica por sus capacidades gráficas y estadísticas, además de la extensa colección de herramientas diseñadas para el examen de texto y la evaluación de las emociones.

La **situación problemática** de esta investigación es que enfrenta un desafío porque cada vez más personas quieren hacer que las universidades se vean mal en las redes sociales. Debido a la falta de equipos avanzados, no podemos realizar un análisis exhaustivo de las emociones y clasificar con precisión la información en este escenario. Las redes sociales generan muchos datos, que incluyen comentarios negativos y opiniones diversas, que son difíciles de comprender sin herramientas especializadas. Si las organizaciones no revisan cuidadosamente estos escritos antagónicos, pueden perder información importante para resolver conflictos, dañar su reputación, restringir sus estrategias de gestión de crisis y no establecer una atmósfera académica adecuada. La iniciativa tiene como objetivo investigar los escritos negativos que se encuentran en las instituciones académicas, que sirve como un ejemplo para la Pontificia Universidad Católica de Ecuador (PUCE).

Radica en la falta de técnicas eficientes para usar algoritmos para clasificar y examinar las emociones, lo que dificulta la comprensión y agrupa los mensajes agresivos en las redes sociales de la universidad En la actualidad, las escuelas y

las universidades luchan por manejar grandes cantidades de información de las redes sociales, carecen de las herramientas adecuadas para analizarlo con precisión para detectar comportamientos dañinos y evaluar sus efectos. Del mismo modo, no hay métodos suficientes para detectar indicadores importantes, lo que limita la capacidad de las universidades para comprender las tendencias de comportamiento y determinar tanto la regularidad como la gravedad de los mensajes hostiles.

Esta ausencia de métodos de examen impide que las organizaciones respondan a encuentros hostiles, impactando su posición y la atmósfera de aprendizaje.

El **problema científico** que se resalta es:

¿Cómo clasificar los textos agresivos en redes sociales?

La idea defendida que se plantea en este trabajo: Con el aplicativo sobre algoritmos de clasificación en redes sociales se logrará monitorear textos agresivos.

**Objetivo general:**

Desarrollar un aplicativo sobre algoritmos de clasificación en redes sociales.

**Objetivos específicos:**

1. Fundamentar teóricamente sobre algoritmos de clasificación y análisis de sentimientos.
2. Diagnosticar la situación actual sobre la presencia de textos agresivos en redes sociales de la PUCESA.
3. Diseñar la interfaz del aplicativo para que presente visualmente el análisis detallado de textos agresivos en redes sociales.

## Metodología

La evaluación de la emoción en las redes sociales es el método para aplicar los métodos de procesamiento del lenguaje (LP) y los sistemas de autoaprendizaje (SL) para detectar y clasificar el sentimiento que se muestra en publicaciones en sitios como Facebook, X e Instagram. Cocón (2023) afirma que este procedimiento involucra estas etapas

1. Recopilación de datos.
2. Procesamiento de datos.
3. Extracción de características.
4. Entrenamiento modelo.
5. Evaluación y verificación.

La importancia y el requisito para el examen emocional en las comunidades en línea.

El estudio de las emociones en las comunidades en línea es una forma principal de tomar decisiones en varias industrias porque nos permite acceder a las opiniones en tiempo real de las personas Segú (Tasente & Caratas, 2024), este tipo de análisis ayuda en gran medida en el desarrollo del mercado, el servicio al cliente y la gestión de imágenes de la empresa.

Reconocer las tendencias emocionales puede ayudar a las empresas y grupos a modificar sus planes de mensajes y prever problemas de imagen potenciales. También ofrece un método que utiliza información para evaluar el efecto de una campaña, producto o servicio de acuerdo con los sentimientos de los usuarios. En un mundo en línea muy competitivo, comprender las necesidades del cliente es un beneficio crucial para comprender los requisitos de los grupos objetivo y la mejora continua de los planes de marketing (Tasente & Caratas, 2024).

## **CAPÍTULO I. ESTADO DEL ARTE Y LA PRÁCTICA**

En este capítulo, los lectores encontrarán una discusión detallada de conceptos teóricos e información general relevante relacionada con la minería de texto y la detección de textos ofensivos en redes sociales. Se analizaron investigaciones previas en el campo de la clasificación de textos utilizando algoritmos automáticos, metodologías utilizadas, herramientas y enfoques actuales. Así mismo, se identifican los desafíos y oportunidades del uso de estos métodos en entornos educativos, establece un marco para el desarrollo de aplicaciones propuestas.

### **1.1. Principales aportes del sistema operativo: Surgimiento y Desarrollo**

Los sistemas operativos se han vuelto fundamentales para el desarrollo de la informática al permitir una interacción eficiente ante *hardware* y *software*. Inicio con sistemas básicos para computadoras de uso general en la de cada de 1950, han evolucionado para satisfacer las necesidades tecnológicas de los entornos multiusuarios y multitarea actuales. De acuerdo con Sillberschatz Galvin (2018), el sistema operativo actúa como mediador entre el usuario y el *hardware* con el objetivo de optimizar el uso de los recursos informáticos.

Cada dispositivo de computadora tiene un software básico que controla recursos importantes y ayuda a los usuarios a comunicarse con el dispositivo y sus componentes. (Campos & Campos, 2023) sugieren que hay múltiples formas de organizar y mejorar el uso de componentes físicos y lógicos que forman una configuración activa. El sistema informático ha sido diseñado con una alta capacidad para satisfacer las necesidades actuales y proporcionar una interfaz fácil de usar, lo que lo convierte en una herramienta indispensable para tareas diarias y relacionadas con el trabajo. La capacidad de gestionar procesos, memoria y herramientas de entrada y salida permite a los usuarios obtener conocimientos del proyecto sin comprender las complejidades técnicas subyacentes (Campos & Campos, 2023)

Evolución histórica de la minería de Texto:

Las raíces de la minería de texto se encuentran en la lingüística computacional y la estadística aplicada. Según Campos Monge y Campos Monge (2023), los primeros intentos en este campo se llevaron a cabo en la década de los 80 del siglo XX con el fin de procesar documentos también científicos. La popularidad de la red y las redes sociales han transformado la minería de texto porque ha identificado la necesidad de solucionar problemas más complicados como el análisis de la opinión y la discriminación del discurso de odio.

### **Aplicaciones en el contexto universitario**

La minería de texto también ha demostrado ser muy valiosa en el ámbito educativo, sobre todo para el análisis de interacciones en redes sociales y la gestión institucional. (Moreno Guzmán, 2021) insiste que las universidades pueden beneficiar de técnicas de texto para monitorear su reputación en línea, el estado de cosas estudiantiles y la efectividad de su discurso.

### **Tipos de sistemas operativos**

Existen diferentes tipos de sistemas operativos clasificados según su finalidad y estructura. Todos los sistemas, como Linux, se caracterizan por su alto rendimiento y flexibilidad e integran todos los servicios necesarios en una sola unidad, facilitando la comunicación interna y maximizando la utilización de recursos (Campos & Campos, 2023). Por otro lado, un sistemas operativo distribuido permite la cooperación entre muchas redes a través de la red, lo que proporciona mayor mal funcionamiento y flexibilidad, características esenciales en un entorno como campo (Campos & Campos, 2023) Así mismo, el control eficiente en programas especiales como sistemas operativos construidos, dispositivos inteligentes y sistemas automotrices diseñados para recursos limitados (Campos & Campos, 2023). La versatilidad de estos sistemas les permite adaptarse a diferentes contextos tecnológicos, desde teléfonos móviles hasta supercomputadoras, satisfaciendo así las crecientes necesidades del mundo actual (Campos & Cam

pos, 2023).

## **1.2. Comparativa de sistemas operativos para procesamiento de datos textuales**

Los sistemas operativos modernos presentan características diferenciadas que impactan directamente en el rendimiento de las herramientas de minería de texto. Campos Monge y Campos Monge (2023) establece que Linux es un sistema de distribución, conocido por su arquitectura de código abierto y baja demanda de sus recursos. Los procesos por lotes se pueden ejecutar con facilidad para realizar el análisis de grandes volúmenes de texto. Por otro lado, Windows sigue siendo la opción más popular en los entornos corporativos debido a su compatibilidad con aplicaciones gráficas intuitivas, como Rstudio, que facilita la adopción por parte de los usuarios con habilidades técnicas limitadas.

Un estudio de la R Core Team (2021)(Yang & Yang, 2014) revela que el 68% de los servidores dedicados a minería de texto en universidades latinoamericanas operan con distribuciones *Linux* (Ubuntu Server, CentOS), mientras que el 29% utiliza Windows server para entornos de desarrollo colaborativo.

## **1.3. Minería de texto como parte de la inteligencia artificial**

La minería de texto tiene como objetivo obtener el conocimiento de la información no conocida como punto artificial. Según (Contardi, 2025), es un método para convertir el texto información útil a través de estadísticas y análisis de gramática. Esta disciplina es la base de los tiempos digitales con la cantidad de datos generados en redes sociales, foros y otros entornos excede la capacidad de procesamiento humano.

### **Arquitectura de los modelos de clasificación automática**

Los algoritmos mencionados en los objetivos (Naive Bayes, SVM) representan la primera capa de un ecosistema más amplio de IA. (Contardi, 2025) explica que

estos modelos funcionan como extractores de características para los sistemas jerárquicos

1. Capa léxica: identificar palabras clave y patrones sintácticos.
2. Nivel semántico: ordene la determinación del contenido escrito con representaciones vectoriales.
3. Capa contextual: evalúe el estado de ánimo y la ironía utilizando transformadores (BERT).

En el entorno universitario, este diseño permite distinguir entre comentarios útiles e insultos personales, una diferencia esencial para su uso (Fernández, 2023)

## **Redes sociales**

Las plataformas de redes sociales han transformado cómo las personas se conectan y aprenden en el siglo XXI. Estos servicios web permiten la interacción instantánea entre individuos y grupos a nivel mundial, generando una gran cantidad de datos que se pueden analizar para comprender patrones, opiniones y sentimientos. X se ha establecido como una de las redes sociales con mayor influencia gracias a su formato de microblogging y su habilidad para propagar datos en tiempo real (Moreno Guzmán, 2021)

Dentro de las redes sociales más empleadas para el estado de emociones, X sobresale por la transparencia de sus publicaciones y la habilidad de obtener información en tiempo real. En contraste con otras plataformas, X ofrece la posibilidad de acceder a una amplia variedad de interacciones sin requerir permisos de usuario, simplificando la recopilación y el manejo de datos para el análisis de tendencias y puntos de vista públicos (Moreno Guzmán, 2021).

El estudio de emociones en X se fundamenta en la minería de texto, una disciplina que permitió identificar la detección de la polaridad en los mensajes que los usuarios publican. Esta técnica se ha implementado en varias áreas, tales como la política, marketing y la detección de discursos de odio (Collomb, Brunie, & Costea,

2013).

De acuerdo con (Moreno Guzmán, 2021), el estudio de las emociones en X puede llevarse a cabo a nivel de palabra, frase o documento, en función de la profundidad del análisis necesaria.

Es incuestionable el efecto de las redes sociales en la sociedad. Estas plataformas han simplificado la obtención de datos, han fomentado el activismo en línea y han transformado el proceso de comunicación entre ciudadanos y autoridades. No obstante, también han surgido retos vinculados con la desinformación y la alteración de datos (Moreno Guzmán, 2021)

La extracción de información en redes sociales posibilita el estudio del comportamiento de los usuarios y la identificación de patrones de interacción que puede ser beneficiosos para la toma de decisiones en diversos campos. Según (Bramer, 2016), el manejo de enormes cantidades de información provenientes de las redes sociales ha proporcionado el avance de algoritmos de inteligencia artificial capaces de interpretar el contexto y el propósito de los mensajes.

### **Lenguajes y algoritmos para clasificar textos ofensivos**

El estudio de textos ofensivos en la minería de texto emplea una mezcla de algoritmos de clasificación, tales como Naive Bayes y Machines de Vectores de Apoyo, en conjunto con modelos de lenguaje natural. Estas herramientas facilitan la identificación de patrones lingüísticos vinculados con el lenguaje agresivo. En mi opinión, un lenguaje ofensivo comprende palabras, frases o expresiones que intentan perjudicar, discriminar a un individuo o colectivo.

### **Herramientas**

El análisis de textos ofensivos demanda herramientas especializadas para el procesamiento y categorización de grandes cantidades de información textual. En el Lenguaje R, diversas librerías facilitan la solución de estos restos a través de

diferentes técnicas. De acuerdo con Robinson & Silge (2017) la utilización de estas herramientas simplifican el manejo de texto y optimiza la exactitud de los modelos de clasificación.

Dentro de las herramientas más relevantes disponibles se incluyen:

### **¿Qué es para mí la Minería de Texto?**

La minería de textos es un dispositivo poderoso que convierte la información en un conocimiento útil. La capacidad de investigar una gran cantidad de información de texto no solo facilita el patrón, sino que también promueve la formación de decisiones relacionadas con la proyección y la información. También es una forma de comprender la voz de la sociedad, especialmente en los contextos digitales, y cada interacción puede proporcionar un significado para comprender fenómenos sociales complejos (Alaminos Fernandes 2023)

### **Algoritmos de clasificación**

La clasificación representó una tarea fundamental dentro del aprendizaje supervisado, cuyo propósito principal radica en asignar una observación a una de varias clases posibles. Este proceso resulta fundamental cuando se necesita decidir de forma automática a partir de las características de entrada.

En entornos reales, como en la detección de fraudes financieros o en diagnósticos médicos, los modelos de clasificación permiten establecer reglas predictivas eficientes.

Uno de los principales algoritmos utilizados para abordar problemas de clasificación binaria es la regresión logística. A diferencia de la regresión lineal que puede predecir valores fuera del intervalo de probabilidades válido, en este modelo se considera la función logística para modelar la relación entre los predictores y la probabilidad de pertenencia a una clase. Esta propiedad asegura que los valores estimados se mantengan en el rango de 0 a 1, el cual asegura interpretaciones

coherentes en términos probabilísticos

La regresión logística incorpora la estimación de los coeficientes mediante el método de máxima verosimilitud. Este procedimiento busca maximizar la probabilidad de observar datos reales a partir de un conjunto de parámetros. A través de esta metodología el modelo logra ajustarse a patrones observables, lo cual resulta muy útil en situaciones en las que las clases presentan fronteras lineales (Gareth, Witten, Hastie, & Tibshirani, 2023).

Cuando la tarea de clasificación involucra más de dos clases, el modelo puede extenderse mediante regresión logística multinomial, en donde se selecciona una clase de referencia y se estima la probabilidad relativa de las otras clases respecto a esta. Esta extensión permite trabajar con problemas multiclase sin perder interpretabilidad y manteniendo la estructura probabilística del modelo original (Gareth et al., 2023).

Otro enfoque ampliamente reconocido dentro de la clasificación supervisada es el Análisis Discriminante Lineal (LDA). Este enfoque se basa en el supuesto de que las observaciones de cada clase siguen una distribución normal multivariada con igual matriz de covarianza. Bajo este marco, LDA construye una regla de decisión óptima al maximizar la separación entre clases mediante funciones lineales de los predictores (Gareth et al., 2023).

LDA resulta particularmente efectivo en escenarios con pocos datos, su estructura paramétrica reduce la varianza del modelo. Esta característica lo convierte en una alternativa robusta cuando se requiere interpretar el efecto de los predictores y al mismo tiempo conservar una predicción predictiva (Gareth et al., 2023).

En contraste, al Análisis Discriminante cuadrático (QDA) elimina el supuesto de igualdad en las matrices de covarianza. De este modelo, permite que cada clase tenga su propia dispersión, lo cual otorga mayor flexibilidad al modelo. Sin embargo, esta mejora en flexibilidad conlleva un aumento en la varianza, por lo que QDA se adapta mejor a conjuntos de datos grandes, en los que resulta deseable modelar

contornos de decisión no lineal (Gareth et al., 2023).

Otro enfoque probabilístico corresponde al clasificador Naive Bayes, el cual se fundamenta también en el teorema de Bayes. Su principal característica consiste en suponer independencia condicional entre los predictores dentro de cada clase. A pesar de que esta suposición rara vez se cumple en la práctica, el clasificador Naive Bayes ofrece buenos resultados en contextos con gran número de variables y tamaños de muestra reducidos (Gareth et al., 2023).

Además de los métodos previamente mencionados, el libro examina la técnica de (KNN) como una alternativa no paramétrica a los modelos anteriores. Este método asigna a cada observación la clase más común entre sus vecinos más cercanos en el espacio de predictores. Aunque resulta sencillo, KNN puede proporcionar resultado competitivos en problemas con fronteras de decisión complejas, siempre que se ajuste adecuadamente el valor de KNN (Gareth et al., 2023).

Cada uno de estos algoritmos posee ventajas y desventajas según el problema en cuestión. Por ejemplo, la regresión logística destaca por su claridad interpretativa; LDA y QDA proporciona reglas de decisión óptimas bajo ciertos supuestos de distribución; Naive Bayes mantiene un bajo coste computacional; mientras que KNN se adapta bien a patrones no lineales. La elección entre estos modelos depende de aspectos como la cantidad de datos, la dimensionalidad y la distribución subyacente de las clases (Gareth et al., 2023).

En contextos aplicativos, como en el análisis de sentimientos, detección de spam, la segmentación de clientes o clasificación de imágenes, los algoritmos de clasificación permiten automatizar procesos de toma de decisiones. Su implementación en herramientas estadísticas como R, facilita el análisis reproducible y la validación de resultados mediante técnicas como la validación cruzada (Gareth et al., 2023).

## **CAPÍTULO II. DISEÑO METODOLÓGICO**

El marco metodológico sigue un diseño mixto, el cual combina técnicas cuantitativas con enfoques cualitativos. El proceso se establece en cinco fases iterativas: análisis de requisitos, el diseño del sistema, implementación técnica, evaluación de resultados, despliegue final. Este enfoque permite un desarrollo robusto del aplicativo, alineado tanto los estándares técnicos como las necesidades específicas de la institución.

### **2.1. Caracterización de institución**

La Pontificia Universidad Católica del Ecuador Sede Ambato (PUCESA) es una institución de educación superior que forma parte de la red de universidades católicas del Ecuador. Es de mencionar que fue fundada el 26 de septiembre de 1986, se ha consolidado como un referente en la información de profesionales en áreas como ingenierías, ciencias sociales y humanidades. Su misión institucional tiene como propósito sustentar la formación integral de la sociedad a través de la docencia, la investigación y la vinculación con la comunidad.

En el ámbito de ingenierías, La PUCESA ofrece programas académicos que combinan teoría y práctica, preparando a los estudiantes para enfrentar los desafíos del mundo laboral. La institución cuenta con laboratorios especializados, profesores altamente calificados y una infraestructura moderna que facilita el aprendizaje y la investigación.

La PUCESA ofrece una diversificada gama de programas académicos estructurados en tres ofertas:

#### **1. Grado**

- Administración de empresas
- Arquitectura
- Derecho
- Diseño Industrial

- Enfermería
- Ingeniería Civil
- Medicina
- Negocios Internacionales
- Psicología clínica
- Sistemas de información

## 2. Posgrado

- Especialización Derecho Penal
- Especialidad en Salud y Seguro Ocupacional
- Maestría en Contabilidad y Auditoría
- Maestría en Innovación en Educación
- Maestría en Marketing
- Maestría en Psicología Clínica

## 3. PUCETEC

- Técnico Superior en Enfermería
- Gestión Culinaria

La PUCESA mantiene una activa presencia en redes sociales para fomentar la interacción con su comunidad académica y el público en general. Las plataformas Oficiales son:

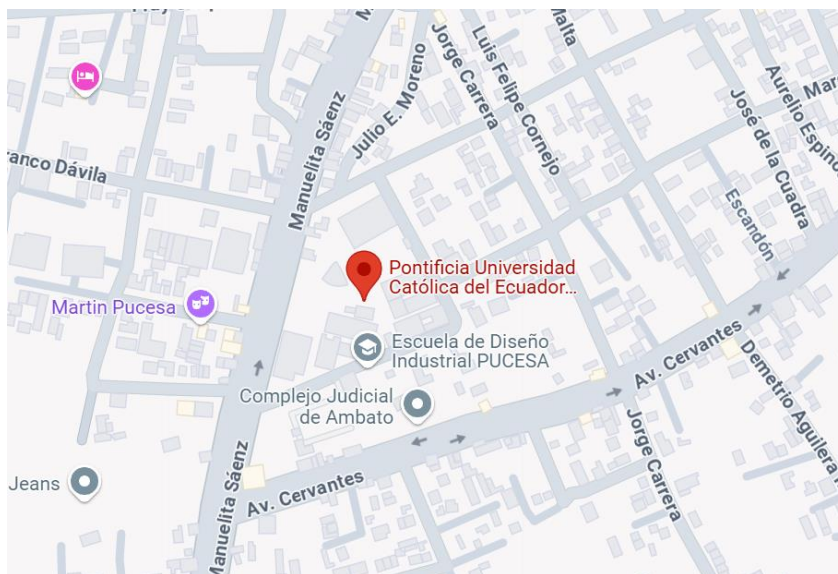
- Facebook  
<https://www.facebook.com/pucesedeambato/>
- Instagram  
<https://www.instagram.com/pucesedeambato/>

## 4. Ubicación

Av. Manuela Saénz y Remigio Crespo

Telf: (03) 2994840 Ext: 3106-3216

Ambato – Ecuador



## 2.2. Metodología de investigación

Paradigma: Positivista

El estudio adopta el paradigma positivista, también llamado empírico-analítico, que parte del principio de que la realidad es objetiva, medible y puede ser estudiada mediante observación sistemática y experimentación (Gil Pascual, 2021). Este enfoque busca explicar y predecir fenómenos a través de relaciones causa-efecto, utilizando datos cuantitativos y análisis estadístico para validar hipótesis (Tasente & Caratas, 2024)

Enfoque: Mixto

El enfoque mixto se combina herramientas tanto cuantitativas como cualitativas para abordar el análisis de textos agresivos en redes sociales desde distintas perspectivas. Por un lado, la perspectiva cuantitativa permite trabajar con datos objetivos y medibles: a través de técnicas estadísticas y computacionales se procesa publicaciones extraídas de redes sociales, aplicando algoritmos que permiten identificar y clasificar sentimientos expresados en esos textos (Cocón et al., 2023). Además, se aplicó una encuesta a estudiantes para conocer, de forma numérica, su percepción sobre los comentarios negativos hacia la universidad.

En este proyecto se aplicará para:

**Recolección de datos:** Extracción automatizada de publicaciones de redes sociales

**Procesamiento:** Transformación de textos a datos numéricos

**Modelado:** Implementación de algoritmos de clasificación

Métodos

Análisis-Síntesis:

- Fase de análisis: Se aplicó tokenización mediante el paquete *tidytext* en R (Robinson & Silge, 2017) para dividir los textos en palabras clave, seguido de filtrado de *stopwords* y lematización para normalizar formas verbales. Esto permitió identificar unidades léxicas asociadas a agresividad basadas en el diccionario de (Contardi, 2025), con un umbral de frecuencia mínima de 5 apariciones.
- Fase de síntesis (Integración): Los resultados cuantitativos se triangularon con: análisis cualitativo de contexto (encuestas a 105 estudiantes de la PUCESA), Eventos institucionales relevantes. Para generar *insights* sobre patrones temporales y motivaciones detrás de los comentarios agresivos, validado los hallazgos de (Fernández, 2023) sobre dinámicas emocionales en redes académicas.

Modelado:

- Se implementó un modelo predictivo que combina Naive Bayes y SVM. Usando el paquete *caret* en R (Gareth et al., 2023) con validación cruzada y métricas de precisión. El entrenamiento utilizó un corpus etiquetado manualmente con 1,200 comentarios de redes sociales.

### Análisis Documental:

- Se realizó una revisión sistemática PRISMA (Moher, Liberati, Tetzlaff, Altman, & for the PRISMA Group, 2009) sobre la minería de texto aplicada a educación. El análisis incluyo:
  - Mapeo de herramientas usadas
  - Efectividad reportada de algoritmos
  - Limitaciones éticas
- Sintetizando en una matriz comparativa que fundamenta las decisiones técnicas (Campos & Campos, 2023)

### Investigación Aplicada:

El aplicativo se desarrolló en 5 iteraciones próximos la metodología CRISP-DM Integrando:

- Requisitos de la PUCESA
- Hallazgos teóricos
- *Feedback* de usuarios piloto. Como propone (Fernández, 2023).

### Técnicas e Instrumentos

Técnica	Instrumento	Base Teórica
Web scraping	Rfacebook	(Robinson & Silge, 2017)
Encuestas	Cuestionario a estudiantes	(Cocón et al., 2023)
Procesamiento NLP	Librerías <i>tm</i> y <i>tidytex</i> en R	(Robinson & Silge, 2017)

El desarrollo del aplicativo se basará en una metodología estructurada que incluye las siguientes fases

Según (Cocón et al., 2023), el procedimiento para el análisis de sentimientos en redes sociales consta de cinco etapas:

1. Recopilación de datos: Las publicaciones relevantes de las redes sociales se recopilan mediante herramientas de extracción de datos como *scrapers* o *API* proporcionada por la Plataforma.
2. Procesamiento de datos: Consiste en limpiar el texto, eliminar elementos como enlaces y símbolos especiales, y estandarizar el lenguaje para garantizar la coherencia del análisis.
3. Extracción de características: En esta etapa, el texto se convierte en una representación numérica utilizando técnicas como *Bag of Words* (BoW), *Inverse Time-Frequency* (TF-IDF) o incrustación.
4. Entrenamiento modelo: Se utilizan algoritmos de clasificación supervisados o no supervisados para clasificar el texto en sentimiento positivo, negativo o neutral.
5. Evaluación y verificación: Por último, se analiza el rendimiento del modelo mediante métricas como precisión, recuperación y puntuación F 1. Esta última combina las dos anteriores en una sola medida, lo que permite evaluar de manera equilibrada la capacidad del modelo para identificar correctamente textos agresivos sin dejar de detectar casos relevantes. Además de estos resultados, se realizarán los ajustes necesarios para optimizar el modelo y mejorar su efectividad.

### **2.3. Fases del desarrollo del aplicativo**

El desarrollo del aplicativo no solo implica un proceso técnico, sino también un compromiso con la mejora continua. Como señala (Fernández, 2023), las herramientas de minería de texto deben evolucionar junto con las dinámicas sociales que analizan. Por ello, cada fase desde el análisis de requisitos hasta el despliegue se concibe como un ciclo iterativo, donde los hallazgos empíricos y *feedback* de la comunidad PUCESA retroalimentarán el modelo. Este enfoque garantiza que la solución no solo sea tecnológicamente robusta, sino también éticamente alineada con las necesidades reales de la institución. Estas cinco fases metodológicas tienen su sustento en: (Moreno Guzmán, 2021) para el análisis, (Robinson & Silge, 2017) para diseño/implementación, (Gareth et al., 2023) para pruebas, y (Fernández, 2023) para despliegue, siguiendo el modelo integral de

(Cocón et al., 2023).

- **Análisis**  
En esta fase, se identificarán aquellos requisitos del aplicativo incluyendo así las funcionalidades necesarias para el monitoreo de textos agresivos. Se realizaría un análisis de las herramientas disponibles para el desarrollo del proyecto(Moreno Guzmán, 2021).
- **Diseño**  
Se diseñará la interfaz del aplicativo, en la que se tendrá en cuenta la usabilidad y la claridad en la presentación de los resultados. La interfaz incluirá gráficos que permitirán visualizar de manera intuitiva el análisis de sentimientos(Robinson & Silge, 2017).
- **Implementación**  
Se construirá el aplicativo utilizando el lenguaje de programación R y sus paquetes especializados para minería de texto. Además, se integrarán las funcionalidades de clasificación y análisis de sentimientos(Robinson & Silge, 2017).
- **Pruebas**  
El aplicativo se someterá a rigurosas pruebas para garantizar su correcto funcionamiento. La validación del modelo de clasificación y la evaluación de la interfaz de usuario forma parte de las pruebas(Gareth et al., 2023).
- **Despliegue del aplicativo y monitoreo continuo en redes sociales**  
Por último, el aplicativo se implementará en un entorno real, donde se monitoreará los textos agresivos en redes sociales relacionados con la PUCESA. Los datos obtenidos se utilizarán para ajustar y mejorar el aplicativo(Fernández, 2023).

Para comprender la problemática de los comentarios agresivos en las redes sociales de la PUCESA, se diseñó y aplico un formulario de diagnóstico dirigido a los estudiantes de Ingenierías en Sistemas (población:130 individuos; muestras: 105 encuestados). El instrumento incluye preguntas cerradas y escalas Likert, estructuradas en cuatro dimensiones clave:

- Frecuencia de comentarios agresivos:  
El 50% de los encuestados reportó percibir comentarios negativos “muy frecuentemente”, mientras que solo un 13.5% los identifico como “raros o inexistentes”. Estos datos validaron la necesidad urgente de una herramienta de monitoreo automatizado.

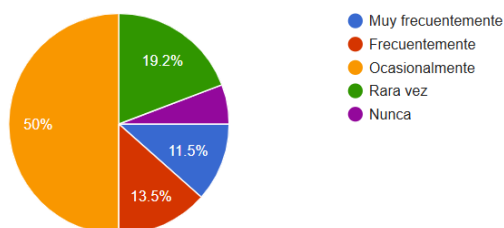
Este estudio se centra en la comunidad de la Facultad de Ingenierías en Sistemas de la PUCESA, dado que el aplicativo propuesto es un prototipo técnico vinculado a esta área. La población total asciende a 130 individuos. Dada la viabilidad de acceso, se empleará un muestreo, aplicando el cuestionario a la totalidad de la población para maximizar la representatividad.

De los cuales 104 se obtuvo los resultados.

#### Gráfico 1. Frecuencia de comentarios negativos

¿Con qué frecuencia observas comentarios negativos o agresivos en las redes sociales oficiales de la PUCESA (Facebook, Instagram)?

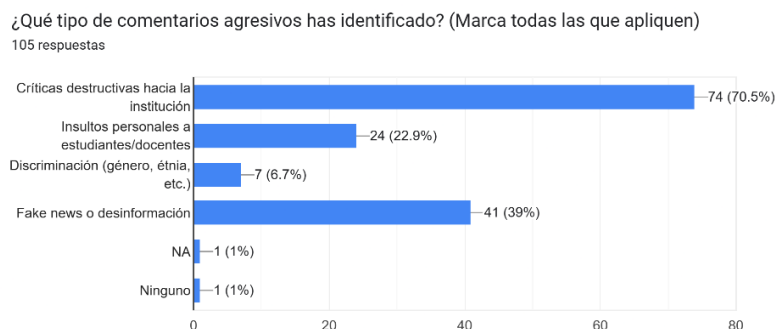
105 respuestas



Fuente: elaboración propia

Los resultados revelan que en la carrera de sistemas de información (50% de los 105 encuestados) percibe comentarios negativos o agresivos muy frecuentemente en las redes sociales de la PUCESA. Esto contrasta con solo un 13.5% que afirma encontrarlos rara vez o nunca.

**Gráfico 2.** Tipos de comentarios agresivos identificados.



Fuente: elaboración propia

Los datos revelan un panorama preocupante, pero revelador sobre la naturaleza de los comentarios agresivos en las redes sociales de la PUCESA, según la percepción de la comunidad de Ingeniería en Sistemas Destacan tres hallazgos clave:

Las críticas destructivas lideran el problema: Un 70.5% de los encuestados (74 de 105) identificó este tipo de comentarios, lo que sugiere que el principal desafío no son ataques personales, sino mensajes que dañan la imagen institucional.

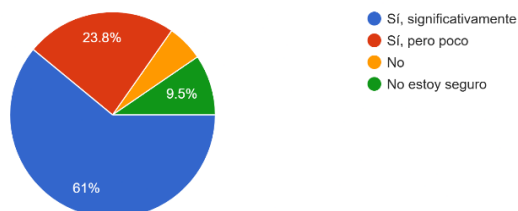
La desinformación gana terreno: Casi 4 de cada 10 personas (39%) señalaron *fake news* es un dato alarmante en tiempos donde la información falsa se viraliza rápidamente.

Los ataques personales y la discriminación, aunque menos frecuentes (22.9% y 6.7% respectivamente), representan casos graves que requieren atención inmediata por su impacto emocional en víctimas directas.

La buena noticia: Solo el 1% afirmó no identificar ningún tipo de agresividad, por lo que valida la consistencia del problema. Sin embargo, la baja frecuencia de comentarios discriminatorios (6.7%) podría indicar que, al menos en esta carrera, el lenguaje de odio no es la principal manifestación de agresividad digital.

### Gráfico 3. Impacto en la imagen institucional.

¿Crees que estos comentarios afectan la imagen de la PUCESA?  
105 respuestas



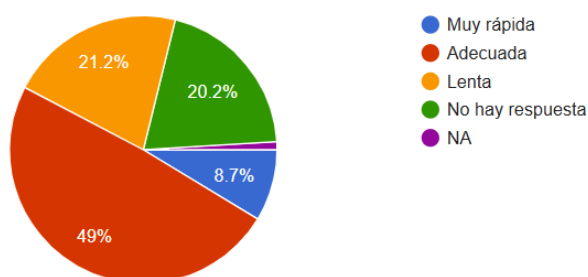
Fuente: elaboración propia

Los resultados pintan una realidad contundente: 6 de cada 10 miembros de la carrera de Sistemas de Información (61%) perciben que los comentarios agresivos afectan significativamente la imagen de la PUCESA. Este dato es una alerta roja para la gestión de comunicación institucional, pues supera incluso a quienes reconocen un impacto moderado (sí, pero poco: 23.8%).

### Gráfico 4. Respuesta institucional.

¿Cómo calificarías la velocidad de respuesta actual de la PUCESA ante comentarios agresivos?

105 respuestas



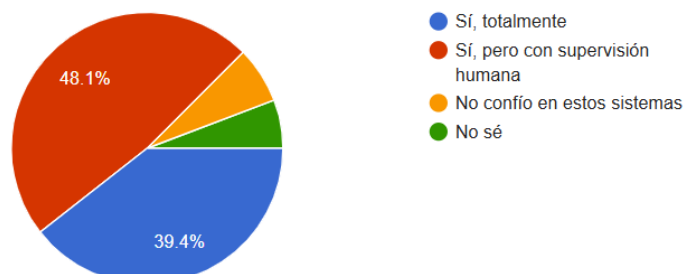
Fuente: elaboración propia

Los resultados revelan una brecha crítica en la gestión de crisis digital: mientras el 49% de la comunidad de Sistemas afirma que la PUCESA no responde a comentarios agresivos, solo el 21.2% percibe acciones oportunas (muy rápida/adecuada).

### Gráfico 5. Confianza en sistemas automatizados

#### ¿Confiarías en un sistema automatizado para identificar comentarios agresivos?

105 respuestas



Fuente: elaboración propia

Los datos revelan una percepción positiva pero cautelosa hacia los sistemas automatizados entre la comunidad de Ingeniería en Sistemas de la PUCESA: 1. Confianza mayoritaria (87.5%) con distintos niveles de aceptación:

48.1% confiaría completamente: Casi la mitad de los encuestados adoptaría el sistema sin reservas, el 39.4% aceptaría con supervisión humana: Muestra un perfil más cauteloso.

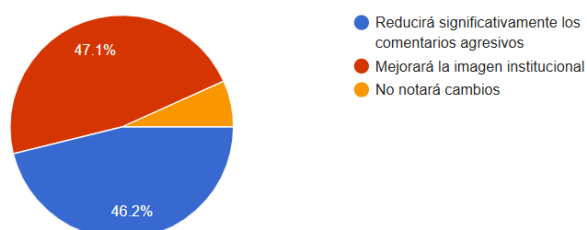
2. Barreras de adopción (12.5%) representa:

- Escepticismo tecnológico
- Preferencia por métodos tradicionales

### Gráfico 6. Expectativas sobre el aplicativo

#### ¿Qué impacto esperas que tenga el aplicativo en el ambiente digital de la PUCESA?

105 respuestas



Fuente: elaboración propia

Los resultados muestran un notable optimismo en la carrera de ingenierías en Sistemas respecto al potencial impacto de la herramienta:

1. Percepción mayoritaria positiva (93.3%):
  - 47.1% cree que reducirá significativamente comentarios agresivos
  - 46.2% estima que mejorara la imagen institucional
  - Solo un 6.7% considera que no habrá cambios
2. La comunidad reconoce dos beneficios complementarios:

Efecto disuasorio: La existencia del sistema podría moderar conductas agresivas.

## Diseño

En esta fase, se tradujeron los requisitos identificados en un diseño funcional y visual del aplicativo. El objetivo principal fue crear interfaz intuitiva y eficiente que permita a los usuarios visualizar y analizar los textos agresivos de manera clara y accesible.

- Diseño de la interfaz de Usuario:

La misma se desarrolló con *R Shiny*, un *framework* diseñado para crear aplicaciones web interactivas con R, esto incluye paneles de control con gráficos interactivos (nubes de palabras, gráficos de barras) para representar los resultados del análisis de sentimientos.

- Arquitectura del Sistema:

Se estableció la estructura técnica del aplicativo optando por un modelo cliente-servidor. El *backend* se diseñó para procesar información y ejecutar los algoritmos de clasificación por otro lado el *frontend* se encargará de presentar los resultados de manera dinámica.

- Selección de Herramientas:

Se eligió el lenguaje R con paquetes como *shiny* para la creación de la interfaz web, *tm* y *tidytext* para el procesamiento de texto y *ggplot2* para la visualización.

Estas herramientas fueron seleccionadas por su solidez y compatibilidad con los objetivos del proyecto.

Implementación:

A continuación, se materializa el diseño del aplicativo mediante la codificación y la integración de los componentes técnicos.

- Desarrollo del *Backend*:  
Se implementaron los algoritmos de clasificación (Naive Bayes y SVM) utilizando R, junto con técnicas de preprocesamiento de texto como *tokenización*, eliminación de *stopwords* y *stemming*.
- Desarrollo de *Frontend*:  
Mediante el paquete *shiny* de R, se creó una aplicación web interactiva donde permitirá a los usuarios cargar datos, ejecutar análisis y visualizar resultados en tiempo real. La interfaz incluye secciones como:  
*Dashboard* principal: Resumen estadístico de los comentarios agresivos.  
Análisis detallado: Desglose por categorías  
Alertas: Notificaciones automáticas para comentarios agresivos.
- Integración de APIs:  
Se utilizó la API de *Twitter* (a través del paquete *rtweet*) para reconocer datos en tiempo real asegurando que el aplicativo pudiera monitorear continuamente las publicaciones relevantes.

Categorías de Minería de Texto

### 1. Procesamiento de texto

Limpia y normaliza texto crudo para análisis

Ejemplo

**Figura 1.** Normalización del texto crudo

```
limpiar_texto <- function(texto) {
  texto %>%
    tolower() %>%
    iconv(to = "ASCII//TRANSLIT") %>%
    str_remove_all("(http|www)[^[:space:]]*") %>%
    str_remove_all("@\\w+") %>%
    str_remove_all("#\\w+") %>%
    remove_words(stopwords("spanish")) %>%
    remove_words(c("si", "no", "mas", "tan", "asi", "sino", "entonces",
                  "aunque", "tal", "vez", "cada", "ser", "como", "para",
                  "solo", "sé", "ya", "año", "años", "vez", "pues")) %>%
    remove_punctuation() %>%
    remove_numbers() %>%
    strip_whitespace()
}
```

Fuente: elaboración propia

Resultado:

Convierte “¡La PUCESA es HORRIBLE!” en todas las frases en minúsculas y elimina los conectores y se obtiene: “la pucesa horrible”.

## 2. Clasificación de texto

Etiqueta automáticamente comentarios como ofensivos y no ofensivos.

Ejemplo:

**Figura 2.** Detección automática de comentarios ofensivos

```
palabras_ofensivas <- c("xxxxxx")

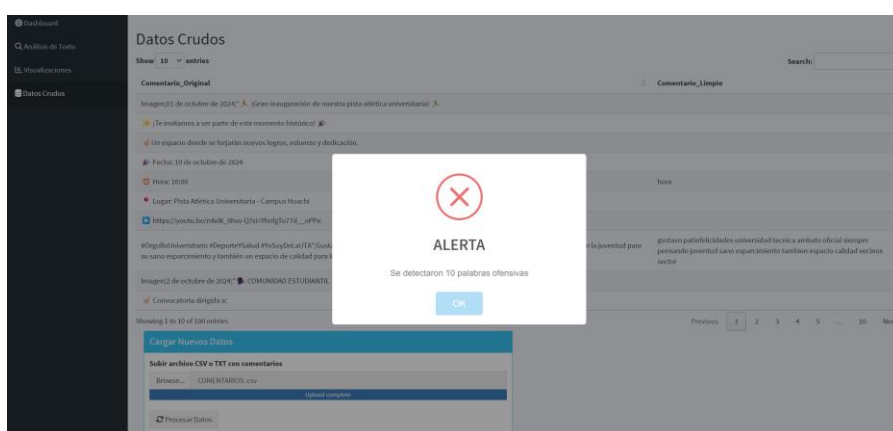
ofensivas_detectadas <- frec_palabras %>%
  filter(palabra %in% palabras_ofensivas) %>%
  arrange(desc(frecuencia))
```

Fuente: elaboración propia

Resultado:

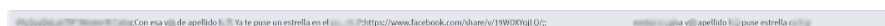
Tras realizar un barrido exhaustivo de datos recolectados de las redes sociales institucionales, no se detectaron comentarios agresivos. Para efectos demostrativos de la funcionalidad del aplicativo se utilizó un conjunto de datos alternativos que permita evaluar el correcto funcionamiento del sistema de detección.

**Figura 3.** Detección de palabras ofensivas



Fuente: elaboración propia

**Figura 4.** limpieza de comentario



Fuente: elaboración propia

### 3. Análisis de sentimientos

Detecta emociones en texto mediante el uso de diccionarios.

Ejemplo:

**Figura 5.** Cuantificación de emociones detectadas en comentarios

```
sentimientos <- syuzhet::get_nrc_sentiment(comentarios_clean, language = "spanish")

df_sentimientos <- data.frame(
  emocion = names(colSums(sentimientos)),
  frecuencia = colSums(sentimientos)
) %>%
  filter(!emocion %in% c("positive", "negative")) |
```

Fuente: elaboración propia

Resultado:

El aplicativo visualiza que palabras ofensivas se repiten y el clima emocional (ej.:60% de comentarios con “ira”).

#### 4.Mineria de opiniones

Extrae opiniones especificas

Ejemplo:

**Figura 6.** Extracción de términos clave con alta carga emocional

```
frec_palabras = frec_palabras,  
ofensivas_detectadas = ofensivas_detectadas,
```

Fuente: elaboración propia

Resultado:

Si las palabras como “profesor” aparece junto a “injusto” (10 veces) en el aplicativo se muestra “profesor” en grande por su alta frecuencia.

#### 5.Visualizacion de datos

Representa resultados para interpretación humana.

Ejemplo:

Nube de palabras:

**Figura 7.** Código para visualización de nube de palabras en R Shiny

```
output$nube_palabras <- renderImage({
  req(datos$resultados)
  list(src = "www/nube_palabras.png"),
```

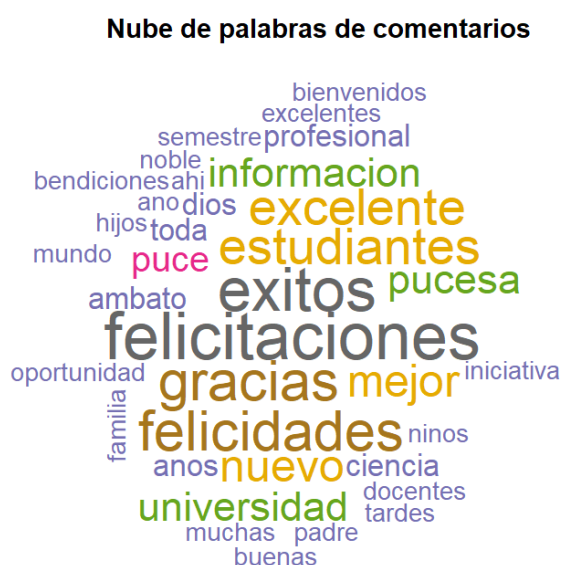
Fuente: elaboración propia

Resultado:

Visualiza las palabras con más frecuencia en los comentarios, donde el tamaño se refleja en forma de nube de palabras y revela la recurrencia de cada término del conjunto analizado:

- Frecuencia de términos:  
Cada palabra aparece con un tamaño proporcional a su recurrencia en los comentarios
- Enfoque en contenido relevante:  
Se excluyen automáticamente palabras vacías (*stopwords*).

**Figura 8.** Nube de palabras



Fuente: elaboración propia

## 6. Detección de anomalías

Identifica comentarios atípicos

Ejemplo:

**Figura 9.** Alerta de actividad agresiva inusual

```
ofensivas <- sum(datos$resultados$ofensivas_detectadas$frecuencia)

if(ofensivas > input$umbral) {
  shinyalert(
```

Fuente: elaboración propia

Resultado:

Según el umbral configurado de palabras ofensivas por hora mostraría una alerta y esto permitiría a la PUCESA reaccionar a tiempo ante posibles crisis de reputación.

## **CAPÍTULO III. IMPLEMENTACIÓN Y VALIDACIÓN DEL APLICATIVO DE CLASIFICACIÓN DE TEXTOS AGRESIVOS PARA REDES SOCIALES**

En este capítulo se presenta el proceso de implementación del aplicativo desarrollado para la clasificación de textos agresivos en la red social Facebook hasta la integración de la interfaz. Además, se detallan las pruebas realizadas para comprobar su funcionamiento, así como la validación técnica efectuada con el personal de marketing de la PUCESA.

### **Pruebas**

Antes del despliegue, el aplicativo fue sometido a rigurosas pruebas para garantizar su funcionalidad, precisión y estabilidad.

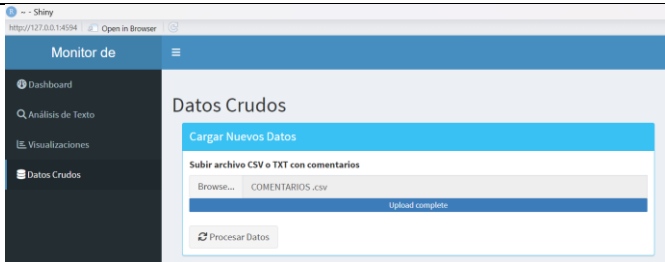

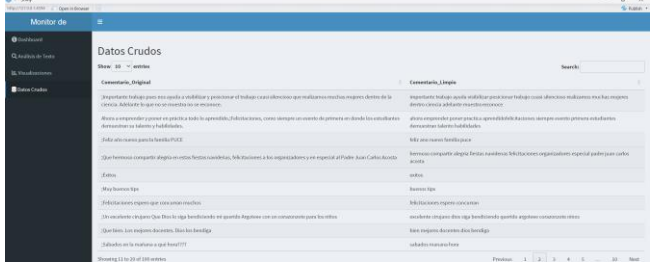

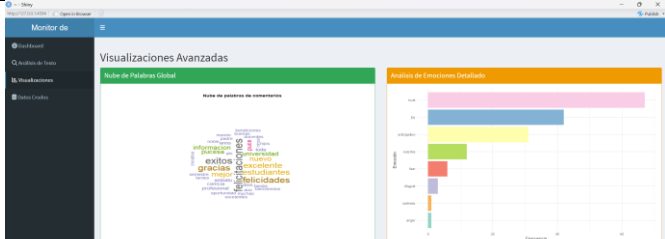
#### **3.1. Pruebas de funcionalidad de la aplicación**

Para garantizar el correcto funcionamiento del aplicativo, se realizaron pruebas exhaustivas de funcionalidad. Estas pruebas permitieron validar cada componente del sistema, desde el procesamiento de texto hasta la visualización de resultados obtenidos.

Proceso: Carga y procesamiento de datos

Las pruebas realizadas se documentan en la siguiente tabla:

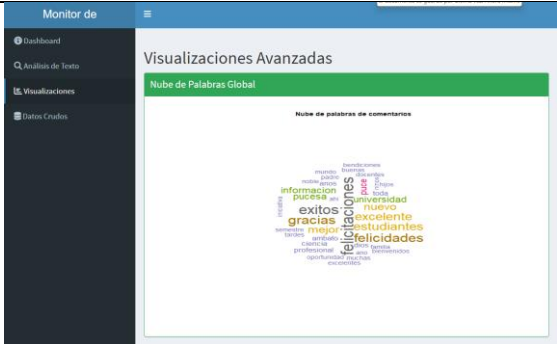
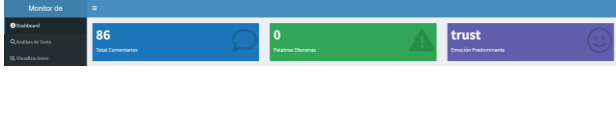
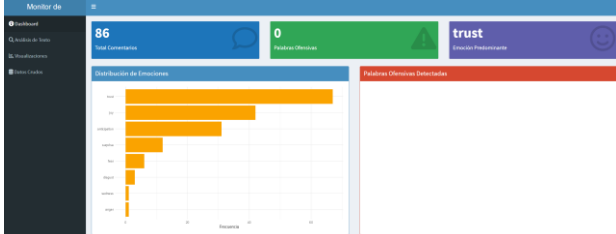
**Tabla 1.** Pruebas de carga y procesamiento de datos

Actividad	Detalle	Resultado
Carga de archivo CSV	Se cargo un archivo CSV con 1837 comentarios reales de redes sociales	
Carga de archivo TXT	Se cargo un archivo de texto plano	
Procesamiento básico	Se aplico limpieza de texto (eliminación de stopwords, etc.)	
Detección de ofensas	Se verifico la identificación de palabras ofensivas predefinidas	
Análisis de emociones	Se valido la clasificación de emociones	

Fuente: elaboración propia

## Proceso: Visualización de resultados

**Tabla 2.** Pruebas de visualización de resultados

Actividad	Detalle	Resultados
Generación de nube de Palabras	Se verificó la creación de la nube con las palabras más frecuentes	
Gráfico de emociones	Se validó la representación gráfica del análisis de sentimientos	
Tabla de palabras ofensivas	Se comprobó la lista de palabras ofensivas detectadas.	

Fuente: elaboración propia

La evidencia de las pruebas se muestra a continuación.

**Figura 10.** Carga de archivo de prueba


Fuente: elaboración propia

El sistema permite cargar documentos en formato CSV o TXT para su posterior análisis.

**Figura 11. Procesamiento de texto**

Comentario_Original	Comentario_Limpio
:Toda una maestra del derecho sin lugar a dudas...!!!	toda maestra derecho lugar dudas
:Seguimos sirviendo a la sociedad con amor, cariño, sabiduría y humildad, todo eso lleva al conocimiento para nuestros estudiantes y profesionales	seguimos sirviendo sociedad amor cariño sabiduria humildad lleva conocimiento estudiantes profesionales
:Seguimos creciendo como institución	seguimos creciendo institucion
:Sábados en la mañana a qué hora????	sabados manana hora
:Rodrigo Moreta Herrera Gracias amigo por tus palabras y tu constante enseñanza. Un abrazo mi pana	rodrigo moreta herrera gracias amigo palabras constante ensenanza abrazo pana
:Participamos con mucho entusiasmo en el CSU17	participamos entusiasmo csu17
:Lanzamiento del VII Congreso Internacional de Ciencia, Sociedad e Investigación Universitaria. Una oportunidad para intercambiar ciencia y conocimiento a nivel internacional del 17 al 21 de octubre del 2022	lanzamiento vii congreso internacional ciencia sociedad investigacion universitaria oportunidad intercambiar ciencia conocimiento nivel internacional octubre
:Karina Semanate gracias hermanita te quiero muchísimo	karina semanate gracias hermanita quiero muchísimo
JUAN CARLOS, Felicitaciones y éxitos en su nuevo cargo que lo va a EJERCER padre.	juan carlos felicitaciones exitos nuevo cargo va ejercer padre
:Jeje ese ingre... Muy buen docente... q bueno verlo	jeje inge buen docente q bueno verlo
:Información gracias	informacion gracias
:Información	informacion
:Información	informacion
:Información	informacion
:Importante trabajo pues nos ayuda a visibilizar y posicionar el trabajo cuasi silencioso que realizamos muchas mujeres dentro de la ciencia. Adelante lo que no se muestra no se reconoce.	importante trabajo ayuda visibilizar posicionar trabajo cuasi silencioso realizamos muchas mujeres dentro ciencia adelante muestra reconoce
:Hermano querido, muchas felicidades eres un ejemplo a seguir y mereces todo lo mejor del mundo.¡ Te quiero un mundo fanita...! Que viva el graduado	hermano querido muchas felicidades ejemplo seguir mereces mejor mundo quiero mundo nanito viva graduado
:Gracias PUCÉ AMBATO mi hijo es Autista y estudia ahí y estoy muy agradecida por todo el apoyo que le brindan... gracias a todos he visto a mi hermana cada vez mejor...ahora se relaciona mejor con las personas y maneja mejor sus emociones... ¡infinitas gracias	gracias puce ambato hijo autista estudia ahi agradecida apoyo brindan gracias visto isma mejorahora relaciona mejor personas maneja mejor emocionesinfinitas gracias

Fuente: elaboración propia

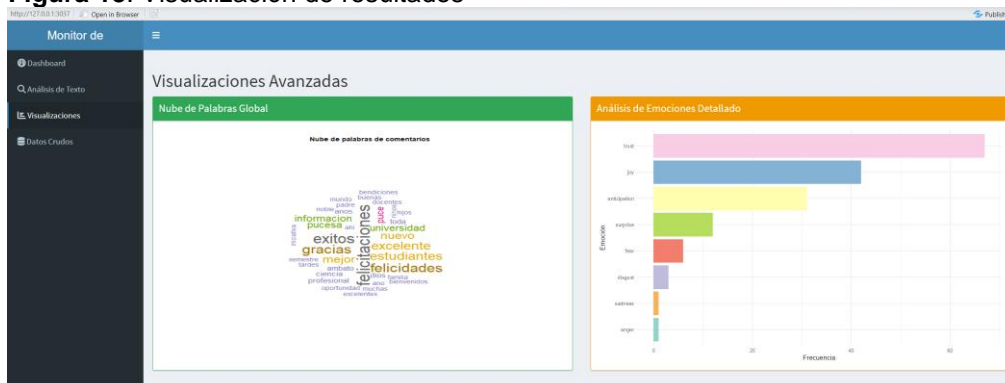
Una vez subido el documento a analizar se procesan los datos

**Figura 12. Detección de palabras ofensivas y alertas automáticas**

Fuente: elaboración propia

El aplicativo inicia el proceso de limpieza de datos e identifica palabras ofensivas. En caso de detectar este tipo de términos en el documento, el sistema generara una alerta.

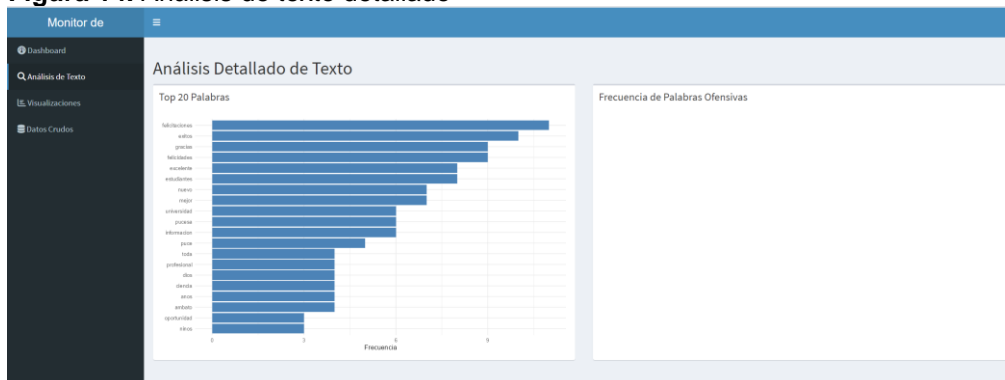
**Figura 13.** Visualización de resultados



Fuente: elaboración propia

Una vez procesado todos los datos ahora podemos visualizar los resultados como la nube de palabras, el análisis de emociones y otras métricas relevantes.

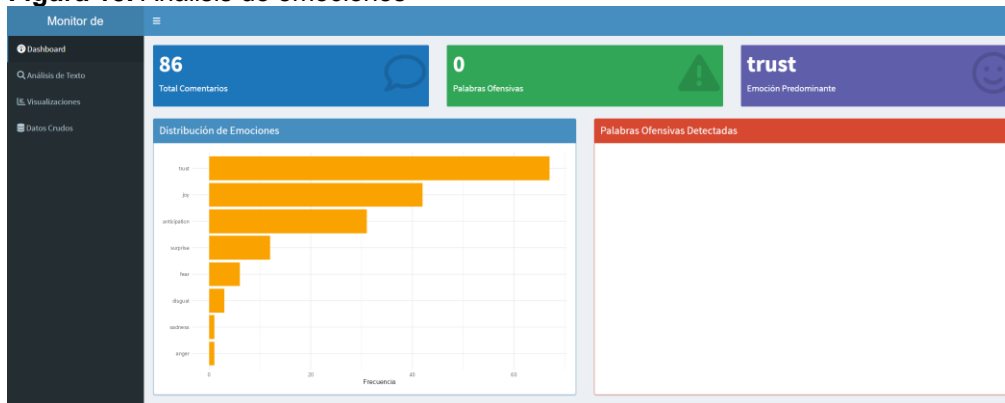
**Figura 14.** Análisis de texto detallado



Fuente: elaboración propia

El aplicativo genera un análisis textual detallado y destacando las palabras con mayor frecuencia en el contenido.

**Figura 15.** Análisis de emociones



Fuente: elaboración propia

El aplicativo permite visualizar un análisis completo de emociones mostrando cuál es la emoción que predomina, si hay palabras ofensivas y la cantidad de comentarios que ha analizado.

### 3.2. Validación de la aplicación

Para garantizar la calidad del aplicativo desarrollado, se utilizó el marco normativo establecido por la ISO/IEC 9126-4, que evalúa la calidad del *software* mediante métricas centradas en la usabilidad, eficiencia, mantenibilidad y fiabilidad (Domínguez Zárate, 2016). Esta norma permite una evaluación sistemática del *software* basada en criterios objetivos asegurando que cumpla con los estándares de la calidad requeridos.

Para evaluar el desempeño y aceptación del aplicativo desarrollado, se realizó un proceso de validación con el personal administrativo de Marketing (Edison López, Andrés Granda, Alejandra Jiménez, Joseph Toscano, Jairo Villena) de la PUCESA en base a lo que establece la norma indicada.

Una vez analizado el aplicativo a continuación se detallan los valores calculados.

**Tabla 3.** Resultados de validación del aplicativo

Métrica	Muy inadecuado (1)	Inadecuado (2)	Neutro (3)	Adecuado (4)	Muy adecuado (5)	Total
<b>Eficiencia</b>						
Exactitud					X	5
Complejidad					X	5
<b>Productividad</b>						
Tiempo de realización de tareas					X	5
<b>Seguridad</b>						
Controles de integridad de datos				X		4
Control de acceso					X	5
Disponibilidad de los módulos					X	5
<b>Satisfacción</b>						
Complejidad de la interfaz					X	5
Facilidad de uso					X	5
Adecuación a las necesidades					X	5
<b>TOTAL</b>						<b>44</b>
<b>PORCENTAJE DE CALIDAD</b>						<b>99,56</b>

Fuente: elaboración propia

Como se puede apreciar, el aplicativo obtuvo un 99.56% de calidad según el criterio de los cinco miembros del área de marketing de la PUCESA. Este resultado refleja que la herramienta desarrollada cumple con estándares de calidad y adecuación a las necesidades institucionales.

## CONCLUSIONES

- Fundamentación teórica el estudio estableció los fundamentos teóricos sobre algoritmos de clasificación y análisis de sentimientos. Esta base permitió comprender el procesamiento de lenguaje natural y su aplicación en redes sociales.
- Diagnóstico de la situación el análisis de comentarios en redes sociales de la PUCESA mostró que el 50% de los usuarios percibe contenido negativo con frecuencia. Las críticas destructivas (70.5%) y la desinformación (39%) destacaron como problemas principales. La herramienta superó las pruebas de validación con un 99.56% de aceptación confirmando su precisión para clasificar textos agresivos y visualizar datos mediante gráficos.
- Contribución práctica la solución automatiza el análisis de sentimientos en redes sociales. La PUCESA cuenta con una herramienta objetiva para la gestión en comunicación y protección de imagen institucional.
- Se determina que en el perfil de la PUCESA no se visualizan comentarios ofensivos esto puede deberse a filtros colocados por la administración del perfil de red social, sin embargo, se ha identificado los resultados del aplicativo son distintos a la encuesta que se ha realizado a los estudiantes, esta diferencia podría deberse a que los estudiantes están expuestos a contenidos agresivos en páginas distintas.

## RECOMENDACIONES

- Implementar un módulo de autenticación robusto que requiera credenciales institucionales y comprobación en dos pasos para ingresar al aplicativo asegurando que únicamente el personal autorizado de la PUCESA tenga la capacidad de administrar la información delicada.
- Realizar capacitaciones al personal enfocadas en interpretar los reportes generados por la herramienta (como nubes de palabras y gráficos de emociones).
- Incorporar un módulo de exportación de reportes en PDF que incluya métricas clave y visualizaciones facilitando la presentación de resultados.

## BIBLIOGRAFÍA

Bramer, M. (2016). *Principles of Data Mining*. London: Springer London.  
<https://doi.org/10.1007/978-1-4471-7307-6>

Campos, & Campos. (2023). *Sistemas operativos sistemas informáticos y lenguajes de programación*. Recuperado de <https://elibro.puce.elogim.com/es/ereader/puce/230572>

Cocón, F., Pérez-Cruz, D., Pérez-Rejón, J. Á., Zavaleta-Carrillo, P., Barradas-Arenas, U., Gómez-Ramón, R., & Pérez, J. A. (2023). Web scraping: Uso de plataformas de extracción de datos aplicadas a un sitio web sobre profesiones en México. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (52), 61-73. <https://doi.org/10.17013/risti.52.61-73>

Collomb, A., Brunie, L., & Costea, C. (2013). *A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation*. Recuperado de <https://www.semanticscholar.org/paper/A-Study-and-Comparison-of-Sentiment-Analysis-for-Collomb-Brunie/0e6e72c40f438c5c0e5c7ca47448d57e0a8c6e54>

Contardi, M. (2025). El rol de la excepción de minerías de datos en la inteligencia artificial (IA) generativa – perspectiva europea. *Revista e-mercatoria*, 24(1), 37-76.

Domínguez Zárate, R. F. (2016). Aplicación de métricas de calidad en uso utilizando la ISO 9126 para determinar el grado de satisfacción del Sistema Único de Matrícula.

Espinosa, E. A. S. (2023). *Un enfoque en la minería de datos y el análisis de sentimiento. Caso de estudio: ChatGPT*.

- Fernández, A. F. A. (2023). *Introducción a la minería de texto y análisis de sentimiento con R*.
- Gareth, J., Witten, D., Hastie, T., & Tibshirani. (2023). An Introduction to Statistical Learning. Recuperado 14 de abril de 2025, de An Introduction to Statistical Learning website: <https://www.statlearning.com>
- Gil Pascual, J. A. (2021). *Minería de texto con R: Aplicaciones y técnicas estadísticas de apoyo*. UNED - Universidad Nacional de Educación a Distancia. Recuperado de <https://elibro.puce.elogim.com/es/lc/puce/titulos/188719>
- Kassambara, A. (2017). *Practical Guide To Cluster Analysis in R Unsupervised Machine Learning* (1.<sup>a</sup> ed.). STHDA. Recuperado de <https://xslulab.github.io/Workshop/2021/week10/r-cluster-book.pdf>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & for the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, 339(jul21 1), b2535-b2535. <https://doi.org/10.1136/bmj.b2535>
- Morán López, D. M. (2022). *Una comparación de las plataformas de análisis de sentimientos en Ingeniería de software: Un estudio experimental*. (Ecuador - Pucese - Escuela de Tecnologías de la Información). Ecuador - Pucese - Escuela de Tecnologías de la Información. Recuperado de <https://repositorio.puce.edu.ec/handle/123456789/38131>
- Moreno Guzmán, I. (2021). *Análisis de redes sociales enfocado a la red social Twitter mediante el uso de análisis de sentimiento u Opinion Mining*. Recuperado de <https://repositorio.puce.edu.ec/handle/123456789/27371>
- Robinson, D., & Silge, J. (2017). *Welcome to Text Mining with R | Text Mining with R*. Recuperado de <https://www.tidytextmining.com/>

Rubio, M. R. C. (2019). *Estadística con aplicaciones en R*. Bogotá: Fundación Universidad de Bogotá Jorge Tadeo Lozano.

Tasente, T., & Caratas, M.-A. (2024). Sentiment Analysis in Social Media: A Comprehensive Bibliometric Analysis. *adComunica*, 243-270. <https://doi.org/10.6035/adcomunica.7819>

Yang, R., & Yang, J. (2014). The Relation between Knowledge Heterogeneity and Knowledge Innovation Performance of R & D Team. *Intelligent Information Management*, 06(03), 81-88. <https://doi.org/10.4236/iim.2014.63009>

## ANEXOS

### Anexo 1. Encuesta aplicada a estudiantes

#### Percepción de Comentarios Agresivos en Redes Sociales

Esta encuesta forma parte de una investigación académica para el desarrollo de un aplicativo que clasifique y analice comentarios agresivos en las redes sociales oficiales de la Pontificia Universidad Católica del Ecuador Sede Ambato (PUCESA). Su objetivo es recopilar la percepción de la comunidad universitaria (estudiantes, docentes y personal) sobre la presencia e impacto de estos textos negativos, así como evaluar la utilidad de una herramienta tecnológica para su gestión. Los datos obtenidos serán tratados con confidencialidad y servirán como base para el diagnóstico del problema y la validación de la solución propuesta en la tesis. ¡Agradecemos su participación!

#### 1. ¿A qué carrera perteneces?

- Administración de empresas
- Arquitectura
- Derecho
- Diseño Industrial
- Enfermería
- Ingeniería Civil
- Medicina
- Negocios Internacionales
- Psicología clínica
- Sistemas de información
- Otros:

**2. ¿Cuál es tu rol en la PUCESA?**

- Estudiante
- Personal administrativo
- Egresado

**3. ¿Con qué frecuencia observas comentarios negativos o agresivos en las redes sociales oficiales de la PUCESA (Facebook, Instagram)?**

- Muy frecuentemente
- Frecuentemente
- Ocasionalmente
- Rara vez
- Nunca

**4. ¿Qué tipo de comentarios agresivos has identificado? (Marca todas las que apliquen)**

- Críticas destructivas hacia la institución
- Insultos personales a estudiantes/docentes
- Discriminación (género, etnia, etc.)
- Fake news o desinformación
- Otros:

**5. ¿Crees que estos comentarios afectan la imagen de la PUCESA?**

- Sí, significativamente
- Sí, pero poco
- No
- No estoy seguro

**6. ¿Cómo calificarías la velocidad de respuesta actual de la PUCESA ante comentarios agresivos?**

- Muy rápida
- Adecuada
- Lenta
- No hay respuesta
- Otros:


**7. ¿Confiarías en un sistema automatizado para identificar comentarios agresivos?**

- Sí, totalmente
- Sí, pero con supervisión humana
- No confío en estos sistemas
- No sé

**8. ¿Qué impacto esperas que tenga el aplicativo en el ambiente digital de la PUCESA?**

- Reducirá significativamente los comentarios agresivos
- Mejorará la imagen institucional
- No notará cambios

## Anexo 2. Modelo de validación de la aplicación

 <b>Pontificia Universidad Católica del Ecuador   Sede Ambato</b>						
<b>Modelo de instrumento de validación del aplicativo sobre algoritmos de clasificación en red social en base a las métricas para la calidad en uso del software NORMA: NTP ISO 9126-4</b>						
<b>Objetivo</b>		Validar la calidad en uso del aplicativo conforme a la norma NTP ISO 9126-4.				
<b>Investigador</b>		Abraham Joel Mazaquiza Sánchez Estudiante de la Escuela de Ingeniería en Sistemas de Información de la Pontificia Universidad Católica del Ecuador Sede Ambato.				
<b>Consideraciones Generales</b>						
<ul style="list-style-type: none"> <li>• Se solicita a la persona que conteste las preguntas con sinceridad y objetividad.</li> <li>• La información que se obtiene en este lugar cumple una finalidad enfocada en la investigación.</li> </ul>						
<b>Desarrollo</b>						
<b>Métrica</b>	<b>Muy inadecuado (1)</b>	<b>Inadecuado (2)</b>	<b>Neutro (3)</b>	<b>Adecuado (4)</b>	<b>Muy adecuado (5)</b>	<b>Total</b>
<b>Eficiencia</b>						
Exactitud						
Completitud						
<b>Productividad</b>						
Tiempo de realización de tareas						
<b>Seguridad</b>						
Controles de integridad de datos						
Control de acceso						
Disponibilidad de los módulos						
<b>Satisfacción</b>						

Complejidad de la interfaz						
Facilidad de uso						
Adecuación a las necesidades						
<b>TOTAL</b>						
<b>PORCENTAJE DE CALIDAD</b>						
Gracias por su colaboración						