

UNIVERSIDAD NACIONAL DEL SANTA
ESCUELA DE POSGRADO
Programa de Doctorado en Ingeniería de Sistemas e
Informática



UNS
E S C U E L A D E
POSGRADO

**Marco de trabajo para analítica de redes sociales basada
en sistemas inteligentes aplicado al Centro de Cómputo de
la Universidad Nacional del Santa**

**Tesis para optar el grado de Doctor en Ingeniería
de Sistemas e Informática**

Autor:

Ms. Manrique Ronceros, Mirko Martin

Código ORCID: 0000-0002-0364-4237

Asesor:

Dr. Urrelo Huiman, Luis Vladimir

Código ORCID: 0000-0003-1523-2640

DNI N° 40010219

Línea de Investigación
Sistemas Inteligentes

Nuevo Chimbote - PERÚ
2025



UNS
ESCUELA DE
POSGRADO

CONSTANCIA DE ASESORAMIENTO DE LA TESIS

Yo, **Dr. Luis Vladimir Urrelo Huiman**, mediante la presente certifico mi asesoramiento de la Tesis de doctorado titulada: “**Marco de trabajo para analítica de redes sociales basada en sistemas inteligentes aplicado al Centro de Cómputo de la Universidad Nacional del Santa**”, elaborado por el Maestro Mirko Martin Manrique Ronceros, para obtener el Grado Académico de Doctor en Ingeniería de Sistemas e Informática en la Escuela de Posgrado de la Universidad Nacional del Santa.

Nuevo Chimbote, Marzo de 2025

Dr. Luis Vladimir Urrelo Huiman

ASESOR

CODIGO ORCID: 0000-0003-1523-2640

DNI N° 40010219



UNS
ESCUELA DE
POSGRADO

CONFORMIDAD DEL JURADO EVALUADOR

“Marco de trabajo para analítica de redes sociales basada en sistemas inteligentes aplicado al Centro de Cómputo de la Universidad Nacional del Santa”

TESIS PARA OPTAR EL GRADO DE DOCTOR EN INGENIERÍA DE SISTEMAS
E INFORMÁTICA

Revisado y Aprobado por el Jurado Evaluador:

Dr. Juan Pablo Sánchez Chávez

PRESIDENTE

CODIGO ORCID: 0000-0002-3521-7037

DNI N° 17808722

Dra. Lizbeth Dora Briones Pereyra

SECRETARIO

CODIGO ORCID: 0000-0003-0626-7227

DNI N° 32960646

Dr. Luis Vladimir Urrelo Huiman

VOCAL

CODIGO ORCID: 0000-0003-1523-2640

DNI N° 40010219



UNS
ESCUELA DE
POSGRADO

ACTA DE EVALUACIÓN DE SUSTENTACIÓN DE TESIS

A día uno del mes de febrero del año 2025, siendo las 10:00 a.m. horas, en el aula P-01 de la Escuela de Posgrado de la Universidad Nacional del Santa, se reunieron los miembros del Jurado Evaluador, designados mediante Resolución Directoral N° 169-2025-EPG-UNS de fecha 19.02.2025, conformado por los docentes: Dr. Juan Pablo Sánchez Chávez (Presidente), Dr. Lizbeth Dora Briones Pereyra (Secretaria) y Dr. Luis Vladimir Urrelo Huiman (Vocal); con la finalidad de evaluar la tesis titulada: **"MARCO DE TRABAJO PARA ANALÍTICA DE REDES SOCIALES BASADA EN SISTEMAS INTELIGENTES APLICADO AL CENTRO DE CÓMPUTO DE LA UNIVERSIDAD NACIONAL DEL SANTA"**; presentado por el tesista **Ms. Mirko Martin Manrique Ronceros**, egresado del programa de Doctorado en Ingeniería de Sistemas e Informática.

Sustentación autorizada mediante Resolución Directoral N° 212-2025-EPG-UNS de fecha 01 de marzo de 2025.

El presidente del jurado autorizó el inicio del acto académico; producido y concluido el acto de sustentación de tesis, los miembros del jurado procedieron a la evaluación respectiva, haciendo una serie de preguntas y recomendaciones al tesista, quien dio respuestas a las interrogantes y observaciones.

El jurado después de deliberar sobre aspectos relacionados con el trabajo, contenido y sustentación del mismo y con las sugerencias pertinentes, declara la sustentación como APROBADO, asignándole la calificación de 19 (EXELENTE).

Siendo las 11:00 a.m. horas del mismo día se da por finalizado el acto académico, firmando la presente acta en señal de conformidad.


Dr. Juan Pablo Sánchez Chávez
Presidente


Dr. Lizbeth Dora Briones Pereyra
Secretaria


Dr. Luis Vladimir Urrelo Huiman
Vocal/Asesor

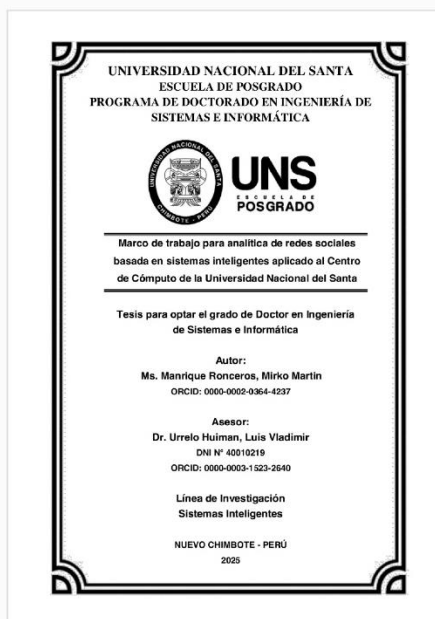


Recibo digital

Este recibo confirma que su trabajo ha sido recibido por **Turnitin**. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega: MIRKO MARTIN MANRIQUE RONCEROS
Título del ejercicio: Informe Final de Doctorado
Título de la entrega: Tesis Doctorado
Nombre del archivo: Tesis_Doctoral_Mirko_Manrique_final.pdf
Tamaño del archivo: 4.08M
Total páginas: 162
Total de palabras: 29,489
Total de caracteres: 177,685
Fecha de entrega: 08-dic-2025 09:47a.m. (UTC-0500)
Identificador de la entrega: 2782666889



Tesis Doctorado

INFORME DE ORIGINALIDAD

19%

INDICE DE SIMILITUD

18%

FUENTES DE INTERNET

3%

PUBLICACIONES

7%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1

repositorio.uns.edu.pe

Fuente de Internet

3%

2

ciencialatina.org

Fuente de Internet

2%

3

www.coursehero.com

Fuente de Internet

1%

4

sedici.unlp.edu.ar

Fuente de Internet

1%

5

Submitted to Universidad Anahuac México
Sur

Trabajo del estudiante

<1%

6

repositorio.puce.edu.ec

Fuente de Internet

<1%

7

repositorio.autonoma.edu.co

Fuente de Internet

<1%

8

www.researchgate.net

Fuente de Internet

<1%

9

redi.unjbg.edu.pe

Fuente de Internet

<1%

10

hdl.handle.net

Fuente de Internet

<1%

repository.udem.edu.co

DEDICATORIA

A Dios, fuente de sabiduría y fortaleza, por guiar mis pasos en cada momento de este largo camino académico y personal. Gracias por iluminarme en las horas de duda, sostenerme en las dificultades y recordarme siempre que todo esfuerzo tiene un propósito más grande.

A mi madre Miriam, ejemplo de entrega, sacrificio y amor incondicional. Su apoyo constante, sus palabras de aliento y su fe en mí han sido el pilar sobre el que he construido este logro. Esta meta alcanzada es también suya.

A mi amada esposa Jenny, compañera de vida, cuyo amor, paciencia y comprensión hicieron posible que perseverara en los momentos más desafiantes. Su compañía ha sido mi mayor refugio y su confianza en mí, la motivación para seguir adelante.

A mi hija Fátima, luz de mis días y motor de mis esperanzas. Este esfuerzo también es un legado para ti, un recordatorio de que con dedicación, constancia y amor todo sueño puede hacerse realidad.

AGRADECIMIENTOS

A mi querida familia, por ser la base sólida que me ha sostenido en cada etapa de este recorrido. Su amor, comprensión y apoyo incondicional me dieron la fortaleza necesaria para perseverar y culminar con éxito este proyecto académico.

A mi asesor Dr. Luis Urrelo Huiman, por su guía constante, su paciencia y su invaluable orientación durante el desarrollo de esta investigación. Sus aportes, sugerencias y exigencia académica no solo enriquecieron este trabajo, sino también mi formación personal y profesional.

A la Universidad Nacional del Santa, mi alma mater; que me abrió sus puertas y me brindó el espacio académico, científico y humano necesario para desarrollar este estudio. Mi gratitud por fomentar un ambiente de crecimiento y excelencia que hizo posible alcanzar esta meta.

A la Escuela de Posgrado de la Universidad Nacional del Santa, por su compromiso con la formación académica de alto nivel y por proporcionar las herramientas y recursos indispensables para la culminación de este doctorado. Su apoyo institucional ha sido fundamental para hacer realidad este logro.

INDICE

RESUMEN	xv
ABSTRACT	xvi
I. INTRODUCCIÓN	1
1.1. Descripción del Problema.....	1
1.2. Formulación del Problema	8
1.3. Delimitación del Estudio	8
1.4. Objetivos	9
1.4.1. Objetivo General	9
1.4.2. Objetivos Específicos.....	10
1.5. Hipótesis.....	10
1.6. Variables.....	10
1.6.1. Variable independiente:	10
1.6.2. Variable dependiente:	10
1.7. Justificación e Importancia.....	10
1.7.1. Justificación Teórica	10
1.7.2. Justificación Práctica	11
1.7.3. Justificación metodológica.....	11
1.7.4. Importancia	12
II. MARCO TEÓRICO	13
2.1. Antecedentes de la investigación.....	13
2.1.1. Internacional	13
2.1.2. Nacional	14
2.2. Fundamentos teóricos de la investigación.....	15
2.2.1. Marco de Trabajo	15
2.2.2. Redes Sociales	16
2.2.2.1. Definición	16
2.2.2.2. Historia y Evolución	16
2.2.2.3. Clasificación	17
2.2.2.4. Métricas en Redes Sociales: Insights, Patrones y Tendencias.....	18
2.2.3. Inteligencia Artificial	18
2.2.4. Machine Learning y Deep Learning	19
2.2.4.1. Machine Learning	19
2.2.4.2. Deep Learning.....	20

2.2.4.3. Modelos supervisados y no supervisados	21
2.2.4.4. Modelos de Deep Learning	21
2.2.5. Procesamiento del Lenguaje Natural (PLN)	23
2.2.5.1. PLN en la clasificación de sentimientos	23
2.2.5.2. Detección de patrones y análisis de tendencias en redes sociales	24
2.2.5.3. Técnicas	24
2.2.6. Análisis de sentimientos	25
2.2.6.1. Modelos de Clasificación en el Análisis de Sentimientos	25
2.2.6.2. Métricas en el Análisis de Sentimientos	27
2.2.7. Sistemas Inteligentes	28
2.2.8. Herramientas y Tecnologías para la Analítica de Redes Sociales	28
2.2.8.1. Bases de datos NoSQL	28
2.2.8.2. PostgreSQL	29
2.2.8.3. Lenguajes y bibliotecas para el análisis de datos	29
2.2.9. Metodología CRISP-DM	30
2.2.10. Gestión	34
2.2.11. Toma de Decisiones	34
III. METODOLOGÍA	35
3.1. Metodología de investigación	35
3.2. Método de investigación	36
3.3. Diseño de investigación	36
3.4. Población	37
3.5. Muestra	37
3.6. Actividades del proceso investigativo	38
3.7. Técnicas e instrumentos de la investigación	39
3.7.1. Técnicas de recolección de datos	39
3.7.2. Instrumentos	40
3.8. Procedimiento para la recolección de datos	40
3.9. Técnicas de procesamiento y análisis de los datos	42
3.9.1. Procesamiento de datos	42
3.9.2. Análisis de datos	43
3.9.3. Validación de resultados	44
3.10. Matriz de Consistencia	44
3.11. Operacionalización de las Variables	46

IV. RESULTADOS Y DISCUSIÓN	47
4.1. Resultados	47
4.1.1. Metodología Propuesta para el Marco de Trabajo de Analítica de Redes Sociales con Sistemas Inteligentes	47
4.1.1.1. Fase 1: Comprensión del Dominio y Definición del Problema (Business Understanding).....	47
4.1.1.2. Fase 2: Comprensión de los datos de Redes Sociales (Data Understanding).....	53
4.1.1.3. Fase 3: Preparación de los datos para el Análisis (Data Preparation) ..	64
4.1.1.4. Fase 4: Análisis Exploratorio de Redes Sociales (EDA - Exploratory Data Analysis)	71
4.1.1.5. Fase 5: Modelado con Sistemas Inteligentes	76
4.1.1.6. Fase 6: Evaluación y Validación de Resultados	80
4.1.1.7. Fase 7: Implementación y Despliegue del Marco de Trabajo	85
4.1.2. Contrastación de la Hipótesis.....	91
4.1.2.1. Dimensión Eficiencia Tecnológica	91
4.1.2.2. Dimensión Patrones y Tendencias	99
4.1.2.3. Dimensión Eficacia Tecnológica	107
4.1.2.4. Dimensión Calidad del Análisis de datos	115
4.2. Discusión	121
V. CONCLUSIONES Y RECOMENDACIONES.....	124
5.1. Conclusiones	124
5.2. Recomendaciones	125
VI. REFERENCIAS BIBLIOGRÁFICAS	126
VII. ANEXOS.....	134
Anexo 01: Cantidad de Alumnos por año	134
Anexo 02: Cantidad de Alumnos por año	135
Anexo 03: Tiempo Promedio de Procesamiento	136
Anexo 04: Cantidad de Insights.....	138
Anexo 05: Precisión en el Análisis.....	140
Anexo 06: Comparación de Modelos de Análisis de Sentimientos	142
Anexo 07: Diagrama de la Metodología CRISP – SNA	146

INDICE FIGURAS

Figura 1: Uso de redes sociales en Perú	3
Figura 2: Uso de redes sociales vs mensajería en Perú en 2023.....	3
Figura 3: Organigrama CECOMP	5
Figura 4: Ciclo de vida de CRISP DM	30
Figura 5: Fase de comprensión del negocio	31
Figura 6: Fase de comprensión de los datos	31
Figura 7: Fase de preparación de los datos	32
Figura 8: Fase de modelado	33
Figura 9: Fase de evaluación	33
Figura 10: Código Python para gráfico de distribución de sentimientos	57
Figura 11: Código Python para limpieza de datos.....	65
Figura 12: Código Python para Preprocesamiento de datos	66
Figura 13: Código Python para conversión de emojis a sentimiento	67
Figura 14: Código Python para almacenar en PostgreSQL:	69
Figura 15: Código en Python para la Distribución de Sentimientos en Comentarios	71
Figura 16: Código Python para Análisis de Palabras Clave y Frecuencia de Términos	72
Figura 17: Código Python para Relación entre Sentimientos y Engagement.....	73
Figura 18: Código Python Identificación de Usuarios Influyentes	75
Figura 19: Código Python para Tendencias de Sentimientos a lo Largo del Tiempo	76
Figura 20: Código Python para Preparación del Conjunto de Datos.....	78
Figura 21: Instalación Hugging Face Transformers	78
Figura 22: Código Python para cargar el Modelo Bert	79
Figura 23: Código Python para Evaluación del Modelo	80
Figura 24: Código Python para evaluación con datos	82
Figura 25: Código Python para Validación en Tiempo Real con Nuevos Comentarios	83
Figura 26: Arquitectura del Sistema	86
Figura 27: Código Python para API REST para Procesamiento en Tiempo Real	87
Figura 28: Código Python para Integración con BD	87
Figura 29: Código Python para Diseño de Dashboard Interactivo	89
Figura 30: Estadísticos Descriptivos Indicador TPP Facebook	92
Figura 31: Estadísticos Descriptivos Indicador TPP Instagram	92
Figura 32: Prueba de Normalidad Indicador TPP Facebook	92

Figura 33: Gráfico Q-Q del indicador TPP Facebook Antes.....	93
Figura 34: Gráfico Q-Q del indicador TPP Facebook Después	93
Figura 35: Prueba de Normalidad Indicador TPP Instagram	94
Figura 36: Gráfico Q-Q del indicador TPP Instagram antes	94
Figura 37: Gráfico Q-Q del indicador TPP Instagram después.....	94
Figura 38: Prueba T Student del indicador TPP Facebook	96
Figura 39: Prueba T Student del indicador TPP Instagram	96
Figura 40: Estadísticos Descriptivos Indicador CIF Facebook	100
Figura 41: Estadísticos Descriptivos Indicador CII Instagram.....	100
Figura 42: Prueba de Normalidad Indicador CIF Facebook.....	100
Figura 43: Gráfico Q-Q del indicador CIF Facebook antes	101
Figura 44: Gráfico Q-Q del indicador CIF Facebook después.....	101
Figura 45: Prueba de Normalidad Indicador CII Instagram	102
Figura 46: Gráfico Q-Q del indicador CII Instagram antes.....	102
Figura 47: Gráfico Q-Q del indicador CII Instagram después	103
Figura 48: Prueba T Student del indicador CIF Facebook	104
Figura 49: Prueba T Student del indicador CII Instagram	104
Figura 50: Estadísticos Descriptivos Indicador PAF Facebook	108
Figura 51: Estadísticos Descriptivos Indicador PAI Instagram.....	108
Figura 52: Prueba de Normalidad Indicador PAF Facebook	108
Figura 53: Gráfico Q-Q del indicador PAF Facebook antes	109
Figura 54: Gráfico Q-Q del indicador PAF Facebook después.....	109
Figura 55: Prueba de Normalidad Indicador PAI Instagram.....	110
Figura 56: Gráfico Q-Q del indicador PAI Instagram.....	110
Figura 57: Gráfico Q-Q del indicador PAI Instagram después	111
Figura 58: Prueba Wilcoxon del indicador PAF Facebook.....	112
Figura 59: Prueba Wilcoxon del indicador PAI Instagram	112
Figura 60: Estadísticas de Prueba del indicador PAF Facebook	112
Figura 61: Estadísticas de Prueba del indicador PAI Instagram	112
Figura 62: Diagrama de Metodología CRISP - SNA	146

INDICE TABLAS

Tabla 1: Operacionalización de las Variables	46
Tabla 2: Indicadores Claves de Rendimiento	50
Tabla 3: Actores claves y responsabilidades	52
Tabla 4: Tipos de datos y formato	54
Tabla 5: Problemas Potenciales en los Datos y Soluciones.....	56
Tabla 6: Almacenamiento de datos	59
Tabla 7: Estrategias para mejorar la representatividad.....	63
Tabla 8: Tareas de limpieza.....	65
Tabla 9: Esquema de base de datos final	69
Tabla 10: Estructura del Dataset.....	77
Tabla 11: Comparación con Métodos Tradicionales	82
Tabla 12: Resultado de TPP para Facebook	98
Tabla 13: Resultado de TPP para Instagram	98
Tabla 14: Resultado de TPP para Facebook	106
Tabla 15: Resultado de TPP para Instagram	106
Tabla 16: Resultado de TPP para Facebook	113
Tabla 17: Resultado de TPP para Instagram	114
Tabla 18: Comparación de modelos de análisis de sentimientos	119
Tabla 19: Cantidad de alumnos por año	134
Tabla 20: Cantidad de grupos por año.....	135
Tabla 21: Tiempo Promedio de Procesamiento (Segundos)	136
Tabla 22: Cantidad de Insights	138
Tabla 23: Precisión en el Análisis	140
Tabla 24: Comparación de Modelos de Análisis de Sentimientos	142

RESUMEN

El crecimiento del uso de redes sociales ha generado una gran cantidad de datos no estructurados que requieren técnicas avanzadas para su análisis. En el contexto de la Universidad Nacional del Santa (UNS), el Centro de Cómputo enfrenta dificultades en la interpretación de datos de redes sociales, lo que impacta la gestión y la toma de decisiones estratégicas. Actualmente, no existe un marco de trabajo estructurado que integre sistemas inteligentes para optimizar el análisis de estas plataformas digitales.

El objetivo de esta investigación fue diseñar e implementar un marco de trabajo para la analítica de redes sociales basado en sistemas inteligentes, con el propósito de mejorar la gestión y la toma de decisiones en el Centro de Cómputo de la UNS. Para ello, se aplicó la metodología CRISP-SNA, una adaptación del modelo CRISP-DM, combinando técnicas de minería de datos, procesamiento del lenguaje natural (PLN) y aprendizaje automático. Se recolectaron y analizaron datos de interacciones en redes sociales mediante modelos avanzados de inteligencia artificial, tales como BERT + LSTM, para la clasificación de sentimientos y detección de patrones de comportamiento.

Los resultados evidenciaron mejoras significativas en la eficiencia del análisis de datos. Se logró una reducción del tiempo promedio de procesamiento en un 87.51% en Facebook y un 86.41% en Instagram. Asimismo, la precisión en el análisis de sentimientos alcanzó un 95.65%, superando en un 15% a los modelos tradicionales como Naïve Bayes y SVM. Además, se observó un incremento del 59.23% en Facebook y del 62.38% en Instagram en la cantidad de insights detectados, lo que confirmó la capacidad del sistema para identificar patrones y tendencias con mayor precisión.

En conclusión, la implementación del marco de trabajo propuesto ha demostrado ser una herramienta efectiva para la analítica de redes sociales en la gestión institucional. La integración de sistemas inteligentes no solo ha optimizado la capacidad de análisis del Centro de Cómputo, sino que también ha facilitado la toma de decisiones basadas en datos, mejorando la comunicación y el impacto estratégico de la universidad en su comunidad digital. Se recomienda la expansión del uso del marco a otras áreas académicas y la actualización constante del modelo para adaptarse a nuevas tendencias tecnológicas en la analítica de redes sociales.

Palabras Claves: Analítica de Redes Sociales, CRISP-SNA, Procesamiento de Lenguaje Natural (PLN), Análisis de Sentimientos, Toma de decisiones, Aprendizaje Profundo, UNS.

ABSTRACT

The growth of social media usage has generated a vast amount of unstructured data that requires advanced techniques for analysis. In the context of the Universidad Nacional del Santa (UNS), the Computing Center faces difficulties in interpreting social media data, impacting management and strategic decision-making. Currently, there is no structured framework that integrates intelligent systems to optimize the analysis of these digital platforms.

The objective of this research was to design and implement a framework for social media analytics based on intelligent systems to improve management and decision-making at the UNS Computing Center. To achieve this, the CRISP-SNA methodology was applied, an adaptation of the CRISP-DM model, combining data mining techniques, natural language processing (NLP), and machine learning. Social media interaction data was collected and analyzed using advanced artificial intelligence models, such as BERT + LSTM, for sentiment classification and behavioral pattern detection.

The results showed significant improvements in data analysis efficiency. The average processing time was reduced by 87.51% on Facebook and 86.41% on Instagram. Likewise, sentiment analysis accuracy reached 95.65%, surpassing traditional models such as Naïve Bayes and SVM by 15%. Additionally, there was an increase of 59.23% on Facebook and 62.38% on Instagram in the number of insights detected, confirming the system's ability to identify patterns and trends with greater precision.

In conclusion, the implementation of the proposed framework has proven to be an effective tool for social media analytics in institutional management. The integration of intelligent systems has not only optimized the Computing Center's analytical capabilities but has also facilitated data-driven decision-making, improving communication and the university's strategic impact on its digital community. It is recommended to expand the use of the framework to other academic areas and to continuously update the model to adapt to new technological trends in social media analytics.

Keywords: Social Media Analytics, CRISP-SNA, Natural Language Processing (NLP), Sentiment Analysis, Decision-Making, Deep Learning, UNS

I. INTRODUCCIÓN

1.1. Descripción del Problema

El uso de sistemas inteligentes aplicados a la analítica de redes sociales ha crecido de forma significativa a nivel global, principalmente debido al incremento exponencial del volumen de datos generados por usuarios en plataformas digitales. Según Ali et al. (2021), los datos provenientes de redes sociales, que incluyen textos, imágenes y videos, representan una fuente rica de información, pero altamente compleja, requiriendo herramientas avanzadas para su análisis. Tecnologías como la inteligencia artificial (IA) y el aprendizaje profundo han demostrado ser efectivas para superar las limitaciones de los enfoques tradicionales al procesar grandes volúmenes de datos no estructurados, proporcionando análisis precisos que benefician sectores como la salud, la educación y el comercio. En el ámbito académico, investigaciones como la de Liu et al. (2021) destacan la implementación de algoritmos de optimización que mejoran la eficiencia y reducen la redundancia en el análisis de datos provenientes de redes sociales.

En el contexto internacional, las redes sociales han evolucionado hasta convertirse en plataformas clave para la interacción humana, generando volúmenes masivos de datos no estructurados que presentan tanto oportunidades como desafíos para su análisis. Con más de 4.9 mil millones de usuarios activos en redes sociales en 2023, equivalentes a más del 60% de la población mundial, estas plataformas se consolidaron como un recurso invaluable para estudiar dinámicas sociales, comportamientos y tendencias globales (Statista, 2023). Sin embargo, la complejidad inherente a los datos generados en estos entornos, que incluyen texto, imágenes y videos, requiere herramientas avanzadas para su procesamiento y análisis, lo que ha llevado al desarrollo e implementación de sistemas inteligentes basados en inteligencia artificial (IA) y aprendizaje profundo (Ali et al., 2021).

El uso de IA en la analítica de redes sociales ha permitido abordar limitaciones de los métodos tradicionales, como la falta de capacidad para extraer patrones significativos en grandes conjuntos de datos no estructurados. Por ejemplo, técnicas como las redes neuronales profundas y los modelos de lenguaje natural han demostrado una eficacia significativa en tareas como el análisis de

sentimientos, la minería de opiniones y la identificación de patrones conductuales en usuarios (Baradaran & Nasiri, 2022). Estas herramientas no solo permiten comprender mejor las interacciones humanas, sino que también optimizan procesos en sectores como la salud, la educación y el comercio. Estudios recientes han demostrado que el uso de IA para analizar datos de redes sociales puede aumentar la precisión en la predicción de tendencias hasta en un 35% en comparación con métodos manuales (Tariq et al., 2021).

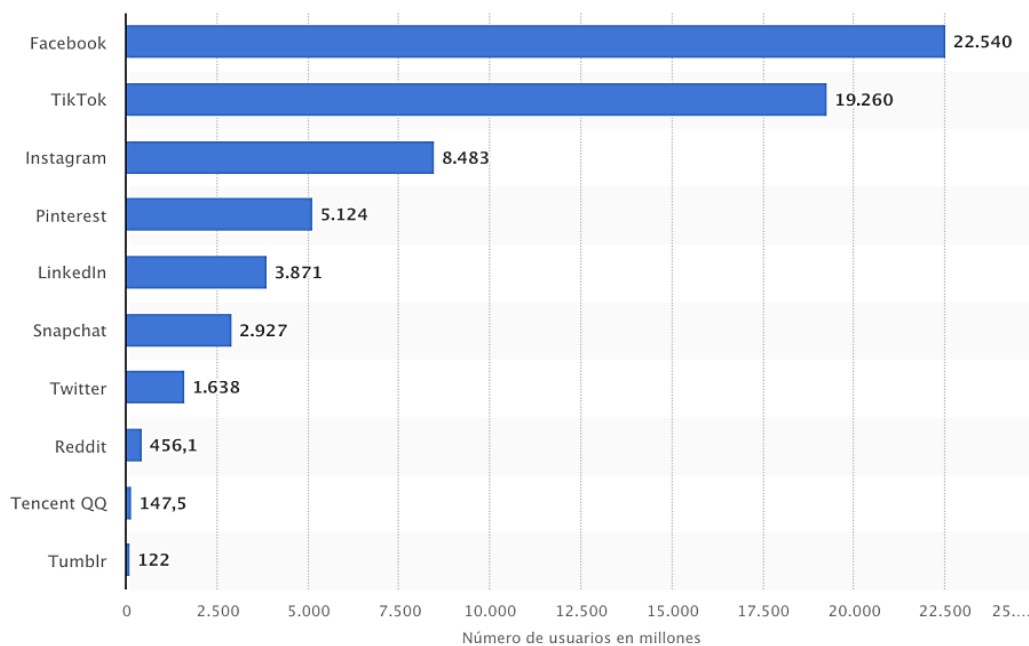
Asimismo, la automatización del análisis de redes sociales ha tenido un impacto transformador en la manera en que las empresas y organizaciones abordan la gestión de datos. Según Kolosiuk y Zinovatna (2024), los sistemas automatizados basados en IA han reducido el tiempo de procesamiento de grandes volúmenes de datos en un 45%, lo que permite a las instituciones responder de manera más efectiva a las necesidades de sus audiencias.

En América Latina, el panorama se caracteriza por una adopción creciente pero desigual de tecnologías avanzadas para la gestión de redes sociales. Según Joshi y Kamath (2023), la región enfrenta desafíos en infraestructura tecnológica y acceso a herramientas avanzadas, limitando la capacidad de análisis a gran escala. Sin embargo, algunos países han dado pasos significativos para integrar sistemas inteligentes en el análisis de redes sociales, enfocados en comprender mejor los patrones de comportamiento y las dinámicas sociales en plataformas como Facebook y Twitter. Esta tendencia es particularmente relevante en el ámbito gubernamental y empresarial, donde la inteligencia artificial se emplea para diseñar estrategias de comunicación y marketing más efectivas (Baradaran & Nasiri, 2022)

En el contexto nacional, Perú ha experimentado un crecimiento acelerado en el uso de redes sociales, con más del 60% de la población activa participando en plataformas digitales como Facebook, Twitter e Instagram. Como lo muestra DataTrust (2025) en sus estadísticas del año 2023, donde se muestra que Facebook sigue siendo el líder en el Perú y en el mundo como red social más usada.

Figura 1:

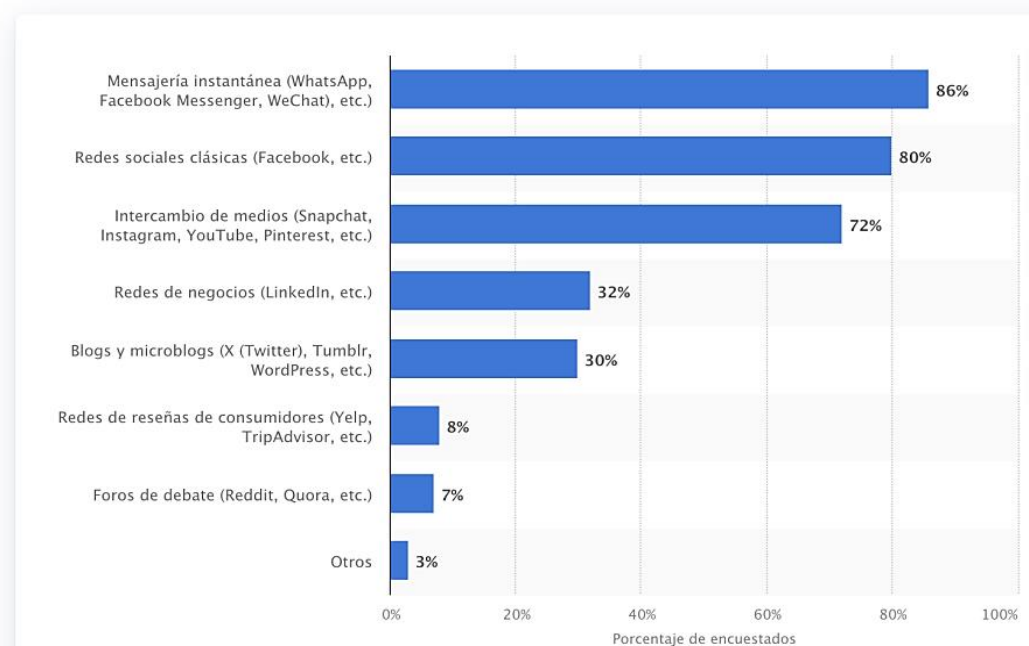
Uso de redes sociales en Perú



Nota: Estadística extraída de <https://es.statista.com/estadisticas/1413665/redes-sociales-con-mas-usuarios-en-peru/>

Figura 2:

Uso de redes sociales vs mensajería en Perú en 2023



Nota: Estadística extraída de <https://es.statista.com/estadisticas/1413665/redes-sociales-con-mas-usuarios-en-peru/>

Sin embargo, el aprovechamiento de estas herramientas en campos como la analítica social aún se encuentra en una etapa incipiente. Este fenómeno responde a limitaciones estructurales como la falta de infraestructura tecnológica avanzada y recursos humanos capacitados en técnicas de inteligencia artificial (IA) y minería de datos, lo cual contrasta con el rápido crecimiento de los datos generados en estos entornos digitales (Vielka Mita et al., 2023). Las iniciativas en IA y sistemas inteligentes aplicados a la analítica de redes sociales son aún limitadas y dispersas. Investigaciones recientes han señalado la importancia de integrar tecnologías avanzadas en contextos específicos como la salud, la educación y la comunicación institucional. Por ejemplo, un estudio realizado por Ponce et al. (2022) mostró cómo algoritmos de aprendizaje automático basados en análisis de sentimientos lograron identificar signos tempranos de problemas de salud mental utilizando datos recopilados de redes sociales. Estos hallazgos resaltaron la potencialidad de los sistemas inteligentes para transformar la manera en que se gestionan y analizan datos en diversos sectores.

Ya existen varias empresas que están utilizando las redes sociales conjuntamente con los sistemas inteligentes para mejorar la experiencia del usuario, como las compañías de teléfonos (Telefónica, Claro), Entidades Financieras (Banco de Crédito, Interbank), Tiendas Online (Linio, Falabella) entre otras. Al combinar las técnicas de Sistemas Inteligentes y analítica de redes sociales, este tipo de empresas son capaces de predecir múltiples acciones como realizar llamadas a clientes en horas disponibles, hacer cálculos de costes y precios, hacer optimizaciones de campañas de marketing o de publicidad en Facebook.

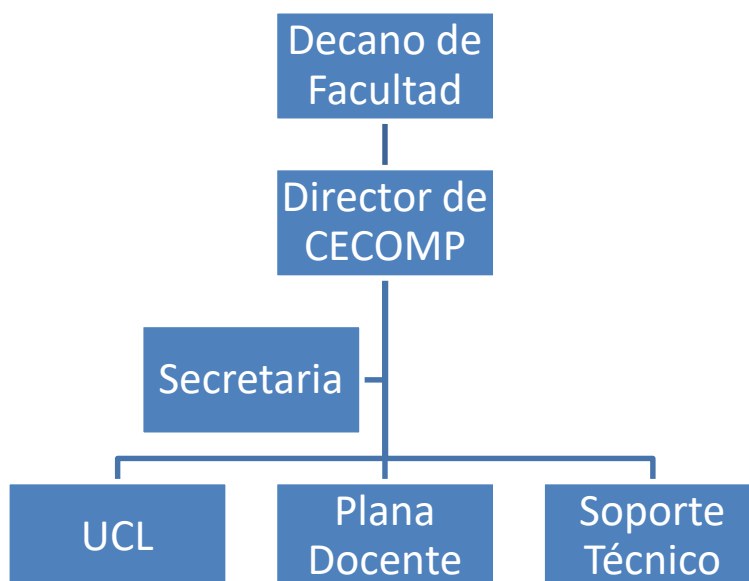
En el ámbito educativo, las universidades en Perú han comenzado a implementar herramientas digitales para mejorar la interacción y comunicación con sus comunidades. Un ejemplo significativo es el desarrollo de sistemas que utilizan algoritmos de optimización genética y análisis de datos para personalizar experiencias turísticas en Tumbes (Sánchez-Ancajima et al., 2024). Este enfoque demuestra cómo las tecnologías avanzadas pueden ser adaptadas a necesidades locales, logrando resultados efectivos en términos de tiempo de respuesta y precisión analítica.

A pesar de estos avances, la mayoría de instituciones académicas y gubernamentales enfrenta retos significativos en la adopción de sistemas inteligentes. Según Castillo-Palacios (2023), la falta de capacitación especializada y la escasa inversión en infraestructura tecnológica han limitado el impacto de estos sistemas en el diseño y ejecución de políticas públicas y estrategias institucionales. Además, estudios como el de Rueda y Cazola (2024) han señalado la necesidad de un marco de trabajo sólido que permita a las organizaciones integrar de manera eficiente las capacidades de análisis y toma de decisiones basadas en datos generados en redes sociales.

En el contexto local, CECOMP es un Centro de Producción perteneciente a la Facultad de Ingeniería de la Universidad Nacional del Santa que brinda cursos extracurriculares orientados a formar especialistas en diversas áreas profesionales y está conformada por:

Figura 3:

Organigrama CECOMP



El centro de Cómputo de la Universidad Nacional del Santa tiene 2700 alumnos (según consulta realizada a base de datos 06 de octubre 2021), en el cual ha ido aumentando en forma considerable a partir del año 2016, según el Anexo 01; permitiendo a la misma vez que se incrementen el número de grupos de estudiantes anualmente en diferentes cursos, según la Anexo 01 y Anexo 02. Esta cantidad de alumnos que han estudiado cursos extracurriculares en el Centro de Cómputo de la

Universidad Nacional del Santa han logrado que se incremente anualmente los ingresos en el centro de producción, según la Anexo 03.

Para poder mejorar la captación, marketing y fidelización de alumnos para el Centro de Cómputo UNS y conozcan nuestros cursos de capacitación que se ofertan, se han tenido que aplicar diferentes estrategias a lo largo de los años como publicidad en afiches, difusión oral a los alumnos de las diferentes escuelas y envío de correo electrónico sin tener mayor éxito o repercusión siendo esto aprovechado por los competidores más cercanos.

CECOMP, desde el año 2016 inició el proceso de Transformación Digital, incorporando en sus procesos de negocio el uso de la tecnología digital cambiando la forma de operar y brindando un valor agregado a sus usuarios. En ese sentido, ha integrado el servicio de página web para mostrar el detalle de los servicios que brinda, sistema de matrícula online de alumnos, sistema de certificados QR, Campus Virtual, sistema de seguimiento administrativo y para mejorar sus estrategias de marketing el uso de redes sociales (Facebook, Twitter, Instagram, Whatsapp) tratando de mejorar la experiencia del usuario en todos los canales de comunicación y ciclo de vida, logrando posicionarse en el mercado regional. teniendo mayor aceptación la red social Facebook que se logrado posicionar en el mercado regional, teniendo en la actualidad 8188 seguidores según Figura 03, que representa un nivel aceptable de público, pero aún insuficiente para liderar el mercado regional.

Las estadísticas actuales de la Fan Page de CECOMP muestran un gran alcance e interacción del público que no se ve reflejado en la cantidad de inscritos en los cursos que demuestran una falta de analítica de la red social que permita recolectar, medir y analizar los datos generados por la actividad de los usuarios en nuestros cursos programados.

Esta falta de un marco de trabajo para realizar una correcta analítica de red social influye en la estrategia, planificación y generación de contenidos, porque no se puede conocer cuáles son los contenidos que mejor funcionan para una determinada audiencia, si se logran convertir nuestros seguidores de las redes sociales en visitantes de nuestra página web y viceversa y cuantos lead o conversiones finales provienen de la red social.

La aceptación y gran demanda de los usuarios en las redes sociales del CECOMP, ha conllevado a visualizar problemas como:

A. Crecimiento exponencial de datos en redes sociales

El uso intensivo de redes sociales por parte de la comunidad universitaria genera un volumen masivo de datos no estructurados, incluyendo publicaciones, interacciones y comentarios. Este flujo constante de datos plantea desafíos relacionados con su procesamiento y análisis, especialmente cuando se busca extraer información relevante para la toma de decisiones estratégicas.

B. Limitaciones en la infraestructura tecnológica y analítica

El Centro de Cómputo de la Universidad Nacional del Santa enfrenta restricciones en términos de infraestructura tecnológica y herramientas especializadas para la analítica de redes sociales. La falta de sistemas inteligentes que integren técnicas de aprendizaje automático, procesamiento del lenguaje natural y análisis de sentimientos limita la capacidad de la institución para monitorear las percepciones de su comunidad y responder de manera oportuna a sus necesidades.

C. Ausencia de un marco de trabajo específico

Actualmente, no existe un marco de trabajo formalizado en la universidad que guíe la implementación de sistemas inteligentes para el análisis de redes sociales. Esto genera un vacío metodológico que dificulta la integración de tecnologías avanzadas en los procesos institucionales, afectando la eficiencia y eficacia en la toma de decisiones basadas en datos.

D. Falta de personal capacitado

El desarrollo y la implementación de sistemas inteligentes requieren de un equipo multidisciplinario con competencias en ciencia de datos, inteligencia artificial y gestión de redes sociales. En el contexto de la Universidad Nacional del Santa, la capacitación en estas áreas es limitada, lo que representa un obstáculo significativo para el éxito de proyectos que involucren tecnologías avanzadas.

E. Impacto limitado en la comunidad universitaria

El uso subóptimo de las redes sociales como herramienta de comunicación y análisis impacta negativamente en la relación entre la universidad y su comunidad. La ausencia de estrategias basadas en datos dificulta la identificación de tendencias, inquietudes y percepciones de los estudiantes, docentes y personal administrativo.

F. Necesidad de decisiones basadas en datos

La universidad requiere decisiones informadas para mejorar sus estrategias de comunicación y fortalecer el vínculo con su comunidad. Sin embargo, la falta de un enfoque sistemático y basado en datos dificulta la optimización de recursos y la mejora continua en los procesos académicos y administrativos.

G. Oportunidades de mejora mediante sistemas inteligentes

La implementación de un marco de trabajo que integre sistemas inteligentes podría optimizar la capacidad del Centro de Cómputo para analizar datos en tiempo real, identificar patrones significativos y predecir comportamientos en redes sociales. Esto no solo mejoraría la gestión institucional, sino que también posicionaría a la universidad como un referente en el uso de tecnologías avanzadas para la educación y la gestión.

En este contexto, resulta fundamental desarrollar un marco de trabajo adaptado a las necesidades y capacidades de las instituciones peruanas, como el Centro de Cómputo de la Universidad Nacional del Santa. La implementación de un sistema de analítica social basado en IA permitiría no solo optimizar los procesos internos, sino también mejorar la interacción con la comunidad universitaria y fortalecer la toma de decisiones estratégicas.

1.2. Formulación del Problema

¿Cómo puede un marco de trabajo para analítica de redes sociales basado en sistemas inteligentes mejorar la gestión y toma de decisiones en el Centro de Cómputo de la Universidad Nacional del Santa?

1.3. Delimitación del Estudio

- El presente estudio se enfocó en el diseño e implementación de un marco de trabajo para la analítica de redes sociales basado en sistemas inteligentes, aplicado específicamente al Centro de Cómputo de la Universidad Nacional del Santa. La investigación se delimitó temporalmente al período comprendido entre enero de 2023 y diciembre de 2024, durante el cual se desarrollaron las etapas de diagnóstico, diseño, implementación y evaluación del marco propuesto.

- En cuanto al alcance geográfico, el estudio se circunscribió a las actividades y datos generados en las redes sociales oficiales administradas por el Centro de Cómputo de la Universidad Nacional del Santa, ubicadas en la ciudad de Chimbote, Perú. Estas redes sociales incluyen plataformas como Facebook e Instagram, utilizadas para la comunicación institucional y la interacción con la comunidad universitaria.
- La población del estudio estuvo constituida por los datos generados por los usuarios de estas redes sociales, tales como publicaciones, comentarios y reacciones, así como por el personal del Centro de Cómputo responsable de la gestión de dichas plataformas. Para la implementación del marco, se emplearon técnicas de minería de datos, procesamiento del lenguaje natural y algoritmos de aprendizaje automático.
- El marco de trabajo se diseñó para abordar problemas específicos como la gestión eficiente del volumen de datos, la identificación de patrones relevantes y la generación de información estratégica para la toma de decisiones. Las variables analizadas incluyeron métricas de interacción en redes sociales, tiempos de respuesta en la gestión de contenidos y la satisfacción de la comunidad universitaria con la comunicación institucional.
- El estudio no incluyó redes sociales no administradas directamente por el Centro de Cómputo ni datos confidenciales de usuarios que no fueran públicos en las plataformas. Asimismo, se limitó a evaluar el impacto del marco en la mejora de la gestión de las redes sociales y la toma de decisiones estratégicas dentro del ámbito institucional, sin extender su alcance a otras áreas de la universidad o entidades externas.

1.4. Objetivos

1.4.1. Objetivo General

Diseñar e implementar un marco de trabajo para la analítica de redes sociales basado en sistemas inteligentes para mejorar la gestión y la toma de decisiones estratégicas en el Centro de Cómputo de la Universidad Nacional del Santa

1.4.2. Objetivos Específicos

- a) Proponer y Diseñar un marco de trabajo de analítica de redes sociales basado en sistemas inteligentes orientado a la experiencia de los usuarios y docentes en el centro de cómputo de la Universidad Nacional del Santa.
- b) Aumentar la eficiencia tecnológica de las redes sociales del centro de cómputo de la Universidad Nacional del Santa.
- c) Elevar la eficacia tecnológica de las redes sociales del centro de cómputo de la Universidad Nacional del Santa.
- d) Incrementar la calidad del análisis de datos de las redes sociales del centro de cómputo de la Universidad Nacional del Santa.
- e) Acrecentar la detección de patrones y tendencias de las redes sociales del centro de cómputo de la Universidad Nacional del Santa.

1.5. Hipótesis

La implementación de un marco de trabajo para analítica de redes sociales basado en sistemas inteligentes mejora significativamente la gestión y la toma de decisiones estratégicas en el Centro de Cómputo de la Universidad Nacional del Santa

1.6. Variables

1.6.1. Variable independiente:

Marco de trabajo para analítica de redes sociales basado en sistemas inteligentes

1.6.2. Variable dependiente:

Gestión y toma de decisiones estratégicas en el Centro de Cómputo de la Universidad Nacional del Santa

1.7. Justificación e Importancia

1.7.1. Justificación Teórica

La investigación se fundamenta en teorías relacionadas con la inteligencia artificial (IA), la minería de datos y el análisis de redes sociales. Desde una perspectiva teórica, este estudio busca contribuir al entendimiento de cómo los sistemas inteligentes pueden optimizar el procesamiento y análisis de grandes volúmenes de datos no estructurados, un desafío ampliamente reconocido en la literatura científica (Ali et al., 2021; Liu et al., 2021). Al integrar técnicas

avanzadas como el aprendizaje automático y el procesamiento del lenguaje natural en un marco de trabajo específico, esta investigación amplía el cuerpo de conocimiento en el campo de la analítica social, proporcionando un modelo replicable para instituciones académicas. Este trabajo también aborda vacíos en la literatura sobre la aplicación de tecnologías inteligentes en contextos locales y educativos, contribuyendo con evidencia empírica y metodológica que fortalezca el desarrollo de nuevas teorías aplicadas.

1.7.2. Justificación Práctica

En el ámbito práctico, esta investigación es de gran relevancia para la gestión institucional en universidades como la Universidad Nacional del Santa, donde las redes sociales juegan un papel crucial en la comunicación con la comunidad universitaria. Actualmente, el Centro de Cómputo de la UNS enfrenta dificultades para gestionar de manera eficiente el creciente volumen de datos generados en redes sociales, lo que limita su capacidad para identificar tendencias, atender necesidades de los usuarios y tomar decisiones estratégicas basadas en datos. La implementación del marco de trabajo propuesto permitirá optimizar la gestión de las redes sociales institucionales, mejorando la interacción con estudiantes, docentes y personal administrativo. Además, los resultados prácticos de esta investigación podrían servir como referencia para otras instituciones académicas que buscan integrar sistemas inteligentes en sus procesos de gestión y comunicación

1.7.3. Justificación metodológica

Desde una perspectiva metodológica, esta tesis doctoral propone el diseño y validación de un marco de trabajo innovador que integra herramientas de inteligencia artificial y técnicas de análisis de datos en tiempo real. El enfoque metodológico utilizado combina el análisis cualitativo y cuantitativo, lo que permite no solo procesar grandes volúmenes de datos, sino también evaluar el impacto del marco en términos de eficiencia y toma de decisiones. Este enfoque metodológico contribuirá a establecer estándares y guías para el desarrollo de proyectos similares en otros contextos, promoviendo la replicabilidad y escalabilidad del modelo propuesto. Además, la investigación aporta un procedimiento sistemático para el diseño, implementación y evaluación de

marcos de trabajo basados en sistemas inteligentes, fortaleciendo las capacidades metodológicas en el campo de la ciencia de datos aplicada

1.7.4. Importancia

Esta investigación es relevante porque aborda problemáticas actuales relacionadas con la gestión de datos masivos en redes sociales, un desafío común en instituciones académicas de países en desarrollo como Perú. Además, su implementación puede transformar la forma en que las universidades utilizan las redes sociales como herramientas estratégicas, potenciando la eficiencia operativa y el impacto institucional. La contribución esperada al conocimiento incluye tanto el desarrollo de un marco teórico y metodológico sólido como la generación de evidencia práctica que pueda ser utilizada para guiar futuras investigaciones y proyectos en el área de la analítica social y los sistemas inteligentes. Con ello, esta tesis no solo beneficiará a la Universidad Nacional del Santa, sino también al ámbito académico y profesional en general, posicionando a la institución como un referente en innovación tecnológica aplicada.

II. MARCO TEÓRICO

2.1. Antecedentes de la investigación

2.1.1. Internacional

Según Ali et al. (2021) tuvieron como objetivo principal desarrollar un marco de monitoreo inteligente para la gestión de datos en redes sociales y sensores portátiles, optimizando la clasificación de datos no estructurados en sectores como la salud. La investigación, publicada en Future Generation Computer Systems, implementó un enfoque basado en técnicas de minería de datos, ontologías y redes neuronales (Bi-LSTM). Los datos fueron procesados en un entorno de computación en la nube y analizados mediante algoritmos de reducción de dimensionalidad y clasificación. Como resultado, se logró una mejora del 30% en la precisión de la predicción de condiciones anómalas de salud y un aumento del 25% en la eficiencia del manejo de datos heterogéneos. Se concluyó que el marco desarrollado es altamente eficiente y adaptable a otros contextos, destacando su potencial para aplicaciones más allá del sector salud.

Según Liu, Song y Liu (2021) tuvieron como objetivo proponer un esquema basado en UAVs para optimizar la trayectoria de recolección de datos en redes sociales mediante tecnologías de inteligencia artificial. La metodología empleada incluyó técnicas de optimización de espacio-tiempo y algoritmos de selección de puntos de muestreo basados en matrices. Los resultados del estudio, publicados en el IEEE Transactions on Network Science and Engineering, demostraron que el esquema propuesto mejoró en un 40% la eficiencia energética al reducir redundancias en la recopilación de datos y aceleró en un 50% el tiempo de procesamiento. La investigación concluyó que la integración de UAVs e IA es una solución viable para maximizar la calidad de los datos recolectados en redes sociales, abriendo nuevas oportunidades para aplicaciones a gran escala.

Según Taherdoost (2023) tuvo como objetivo analizar el uso de algoritmos de aprendizaje automático y redes neuronales para el análisis de redes sociales, enfocándose en la personalización de contenido y la gestión de usuarios. Publicado en la revista Algorithms, este estudio implementó una metodología comparativa que evaluó varios algoritmos de machine learning en redes sociales como Twitter e Instagram. Los resultados indicaron una mejora del 35% en la identificación de patrones de comportamiento y un incremento del 28% en la precisión de las recomendaciones de contenido. El estudio concluyó que la aplicación de redes neuronales y aprendizaje automático en redes sociales no solo optimiza la experiencia del usuario, sino que también ofrece ventajas significativas para empresas y gestores de contenido.

2.1.2. Nacional

Según Ponce, Flores Cruz y Andrade-Arenas (2022) tuvieron como objetivo desarrollar una aplicación basada en aprendizaje automático para la detección temprana de problemas de salud mental en Perú mediante el análisis de redes sociales. El proyecto utilizó la metodología ágil SCRUM y herramientas como Python, SQL Server y Android Studio, combinadas con técnicas de análisis de sentimientos. Los datos fueron recolectados de las plataformas de Facebook y Twitter. Los resultados mostraron que el sistema logró una precisión del 85% en la detección de indicadores de riesgo, lo que representó una mejora del 30% en comparación con métodos manuales. Los autores concluyeron que la integración de herramientas de inteligencia artificial en la analítica social puede tener un impacto significativo en la prevención de problemas de salud mental, mejorando la capacidad de respuesta del sistema de salud en contextos digitales.

Según Sánchez-Ancajima et al. (2024) diseñaron un sistema inteligente para la guía turística en Tumbes, Perú, utilizando un enfoque híbrido basado en reglas y algoritmos genéticos. El objetivo del estudio fue optimizar las rutas turísticas personalizadas según las preferencias del usuario. La metodología incluyó la recopilación de datos georreferenciados y la implementación de técnicas de enrutamiento vehicular. Los resultados indicaron que el algoritmo genético logró una mejora del 45% en la optimización del tiempo de respuesta en

comparación con métodos convencionales. Los investigadores concluyeron que la implementación de sistemas inteligentes adaptados al turismo local puede mejorar significativamente la experiencia del usuario y fomentar el desarrollo turístico sostenible en la región.

Según Huertas Rueda y Linares Cazola (2024) buscaron determinar la influencia de las redes sociales en la competitividad digital de las MSE en el sector ferretero de Chimbote, Perú. Este estudio, de diseño no experimental y nivel correlacional-causal, utilizó cuestionarios estructurados aplicados a 245 consumidores digitales. Los resultados demostraron una relación positiva entre el uso de redes sociales y la competitividad digital, con un incremento del 38% en la percepción de innovación de las empresas que gestionan activamente sus redes sociales. Se concluyó que la implementación estratégica de redes sociales permite mejorar la percepción de modernidad y la participación en el mercado de las MSE locales.

2.2. Fundamentos teóricos de la investigación

2.2.1. Marco de Trabajo

Guerrero y Recaman (2009) “Marco de Trabajo del inglés Framework, se define como un conjunto de componentes físicos y lógicos estructurados de tal forma que permiten ser reutilizados en el diseño y desarrollo de nuevos sistemas de información”.

Los marcos de trabajo, según Guerrero y Suárez (2010) “Contienen patrones y buenas prácticas que apoyan el desarrollo de un producto y un proceso con calidad”.

“Un marco de trabajo (Framework, por su origen en inglés) proporciona un marco de abstracción de una solución para un número de problemas que tienen las mismas similitudes. Generalmente un marco de trabajo esboza o proporciona la idea general de los pasos o fases que se deben seguir en la implementación de una solución sin entrar en los detalles de cuales actividades son realizadas en cada fase.” (Mnkandla, 2009)

2.2.2. Redes Sociales

Las redes sociales, como Facebook e Instagram, han transformado radicalmente la manera en que las personas se comunican e interactúan en el entorno digital. Estas plataformas facilitan la generación en tiempo real de grandes volúmenes de datos no estructurados, lo que posibilita el análisis de comportamientos, la identificación de tendencias emergentes y la comprensión de las dinámicas de interacción a nivel macro. Gracias a esta capacidad, las organizaciones pueden diseñar estrategias de marketing digital y de comunicación basadas en evidencia, optimizando sus procesos y adaptándose a entornos altamente competitivos (Nanda & Kumar, 2021). Las redes sociales generan una enorme cantidad de datos no estructurados, lo que ha llevado a un aumento en el uso de técnicas de inteligencia artificial para analizarlos y extraer patrones de comportamiento (Groshek & Koc-Michalska, 2021).

2.2.2.1. Definición

Las redes sociales son plataformas digitales que permiten la interacción entre individuos mediante la creación de perfiles personales, el intercambio de información y la formación de comunidades en línea (Pérez-Wiesner, Fernández-Martín & López-Muñoz, 2014). Estas plataformas han cambiado significativamente la comunicación interpersonal, facilitando la difusión masiva de contenidos y la generación de redes de contacto a nivel global (Ros-Martín, 2009).

2.2.2.2. Historia y Evolución

El origen de las redes sociales en Internet se remonta a SixDegrees.com, lanzada en 1997, considerada como la primera plataforma de este tipo, ya que permitía a los usuarios crear un perfil y conectarse con otros usuarios de su red de contactos (Boyd & Ellison, 2007). Posteriormente, en el año 2002, surgió Friendster, que introdujo nuevas funciones de interacción social y sirvió como precursor de plataformas más avanzadas (Boyd & Ellison, 2007).

Con el avance de la tecnología, en 2003 apareció MySpace, una red que ganó popularidad entre músicos y artistas debido a su enfoque en

la personalización de perfiles y el uso de contenido multimedia (Ros-Martín, 2009). Sin embargo, fue en 2004 cuando Facebook revolucionó el concepto de redes sociales al enfocarse en conexiones entre individuos y permitir una mayor segmentación de audiencias (Pérez-Wiesner et al., 2014).

El crecimiento de las redes sociales se expandió aún más con la llegada de Twitter en 2006, una plataforma de microblogging que permitió la comunicación en tiempo real mediante publicaciones de 140 caracteres (Ros-Martín, 2009). En 2010, Instagram introdujo un nuevo enfoque visual basado en la publicación de imágenes y videos, convirtiéndose en una de las plataformas más influyentes del mundo digital (Pérez-Wiesner et al., 2014).

2.2.2.3. Clasificación

Las redes sociales pueden clasificarse en dos grandes categorías: horizontales y verticales

- **Redes sociales horizontales o generales:** Estas plataformas no tienen una temática específica y están diseñadas para un público amplio. Ejemplos de estas son Facebook, Twitter y Instagram, donde los usuarios pueden compartir diversos tipos de contenido sin una restricción temática (ONTSI, 2009).
- **Redes sociales verticales o especializadas:** Son aquellas que se enfocan en un nicho o sector específico. Un ejemplo representativo es LinkedIn, orientado a la generación de redes profesionales y el intercambio de información laboral (Pérez-Wiesner et al., 2014). Asimismo, existen plataformas como YouTube, que se centra en la compartición de contenido audiovisual, o Pinterest, que permite la organización de imágenes según intereses temáticos (Ros-Martín, 2009)

2.2.2.4. Métricas en Redes Sociales: Insights, Patrones y Tendencias

Para explotar adecuadamente el valor de la información generada en redes sociales es fundamental contar con métricas que cuantifiquen tanto el alcance como la efectividad de la comunicación digital. Indicadores tales como el número de impresiones, la tasa de engagement, el alcance y la frecuencia de interacción permiten evaluar de manera precisa cómo se difunde el contenido y cuál es el grado de respuesta de la audiencia. Por ejemplo, el alcance indica cuántos usuarios han sido expuestos a un mensaje, mientras que el engagement refleja la interacción activa (likes, comentarios, compartidos) que genera dicho contenido. La combinación de estas métricas ofrece insights valiosos para identificar patrones de consumo y tendencias en el comportamiento del usuario, lo que resulta crucial para ajustar y mejorar las estrategias comunicacionales (González & del Rio, 2019). La tasa de engagement es un indicador clave en la medición del impacto de campañas en redes sociales, permitiendo comprender cómo los usuarios interactúan con el contenido (Hoffmann & Lutz, 2019)

2.2.3. Inteligencia Artificial

La Inteligencia Artificial (IA) es una disciplina científica cuyo desarrollo ha avanzado significativamente en los últimos años, en contraste con otras ciencias que poseen siglos de trayectoria. El término IA se popularizó en 1956, poco después de la Segunda Guerra Mundial. Al igual que la biología molecular, esta área resulta especialmente atractiva para los estudiantes de ciencias, ya que sigue siendo un campo con amplias oportunidades de exploración y descubrimiento en múltiples ámbitos. (Norving, 2008).

Según el Ministerio de Ciencia, Innovación y Universidades (2019), el término "Inteligencia Artificial" se define como la disciplina científica y de ingeniería dedicada al diseño y programación de máquinas capaces de ejecutar tareas que demandan inteligencia para su realización. (p.13).

No obstante, la Inteligencia Artificial no cuenta con una definición única y universal, ya que diferentes autores la abordan desde perspectivas humanas y racionales, lo que ha dado lugar a cuatro enfoques principales y ciertos

desacuerdos. Quienes apoyan el enfoque humano consideran que la IA debe basarse en un método empírico, aplicando hipótesis y validándolas mediante experimentación. Por otro lado, los defensores del enfoque racional proponen una combinación de matemáticas e ingeniería para su desarrollo. (Norving, 2008).

2.2.4. Machine Learning y Deep Learning

2.2.4.1. Machine Learning

El Machine Learning es un enfoque de la inteligencia artificial que permite a los sistemas aprender a partir de datos históricos sin necesidad de una programación explícita. A través de algoritmos que identifican patrones y relaciones, se facilita la automatización de tareas de clasificación, predicción y segmentación en entornos altamente dinámicos. En particular, el aprendizaje supervisado y no supervisado han sido ampliamente utilizados en la construcción de modelos predictivos que mejoran su desempeño a medida que reciben nuevos datos. Gracias a su flexibilidad y capacidad de adaptación, el Machine Learning ha sido aplicado con éxito en sectores como la salud, el comercio electrónico y el análisis de redes sociales (Shalev-Shwartz & Ben-David, 2019).

Este tipo de Inteligencia Artificial identifica patrones a partir de conjuntos de datos o datasets. En el caso del Machine Learning aplicado a la predicción, una de las técnicas más utilizadas es el aprendizaje supervisado, ya que permite aprender a partir de la relación entre características descriptivas o mediante datos de ejemplo e información histórica. Un aspecto destacado es que estos modelos pueden emplearse para predecir nuevas instancias o aplicarse en otros modelos. (Kelleher, Namee, & D'Arcy, 2015)

Chio & Freeman (2018) mencionan que el aprendizaje automático se basa en algoritmos y procesos que aprenden de experiencias y datos previos para prever resultados futuros. En este contexto, el Machine Learning se compone de un conjunto de técnicas aplicadas en sistemas informáticos que, mediante la minería de datos, detectan patrones y elaboran inferencias a partir de la información disponible.

Según Hurwitz & Kirsch (2018), el Machine Learning aprende a medida que se le ingresan datos, sin depender de la programación tradicional. Además, es una herramienta útil para descubrir modelos con mayor precisión a partir del análisis de datos. La solución obtenida permite comprender un prototipo de Machine Learning mediante un algoritmo que se alimenta de los datos capturados. Posteriormente, a través de la incorporación de nuevos datos, se genera un modelo que da lugar a un algoritmo capaz de predecir información futura.

2.2.4.2. Deep Learning

El Deep Learning (DL) es una subdisciplina del Machine Learning (ML) que se distingue por su capacidad de aprendizaje progresivo mediante múltiples capas de representación, cada vez más complejas y abstractas. La profundidad en el aprendizaje profundo hace referencia a la cantidad de capas que componen un modelo de datos, lo que influye en su capacidad de representación. En DL, estas estructuras por capas se implementan mediante modelos conocidos como redes neuronales artificiales, las cuales están organizadas en múltiples niveles superpuestos. En términos generales, el DL proporciona un marco matemático para el aprendizaje de representaciones a partir de los datos (Chollet, 2018, Capítulo 1, p. 8). El Deep Learning (DL) es una técnica dentro del Machine Learning (ML) que utiliza redes neuronales artificiales y se caracteriza por su arquitectura con múltiples capas ocultas (Trusculescu et al., 2020). Según Artola (2019) y Sarmiento (2020), los modelos de DL son ampliamente empleados en tareas de reconocimiento de imágenes.

El Deep Learning se basa en el uso de redes neuronales artificiales de múltiples capas para extraer representaciones jerárquicas y abstractas de la información. A diferencia de los métodos tradicionales, esta técnica permite que los sistemas interpreten datos no estructurados, como imágenes, audio y texto, con niveles de precisión superiores. Modelos basados en redes neuronales convolucionales (CNN) y recurrentes (RNN) han demostrado su eficacia en tareas como el reconocimiento facial, la traducción automática y el análisis de

sentimientos. No obstante, el Deep Learning requiere grandes volúmenes de datos y un alto poder computacional para su entrenamiento, lo que ha impulsado el desarrollo de infraestructuras optimizadas y hardware especializado (Goodfellow, Bengio & Courville, 2019).

2.2.4.3. Modelos supervisados y no supervisados

A. Modelos supervisados

Es el proceso en el que se introducen algoritmos, proporcionándoles preguntas, atributos y respuestas, conocidas como etiquetas, con el propósito de que puedan analizarlas, combinarlas y generar predicciones (Candia, 2019)

B. Modelos no supervisados

Es necesario proporcionarle datos de entrada para que pueda identificar y organizar patrones de comportamiento, permitiéndole así generar un resultado. (García, 2019)

2.2.4.4. Modelos de Deep Learning

A. Redes Neuronales Convolucionales (RNC)

Una RNC es uno de los algoritmos más reconocidos en el Deep Learning, ya que representa un modelo de aprendizaje automático cuyo objetivo principal es la clasificación de imágenes, videos, textos e incluso sonidos. Su capacidad para detectar patrones facilita el análisis automatizado, evitando que los investigadores tengan que extraer características de forma manual (Maeda, 2019).

Las redes neuronales convolucionales presentan tres capas:

- i) **Capa Convolutiva:** Según Voulodimos et al. (2018), la capa convolutiva es una de las principales dentro de una RNC, ya que emplea múltiples núcleos para procesar la imagen completa, así como los mapas de características intermedias, generando diversos mapas de características. Por su parte, Oquab et al. (2015) y Szegedy et al. (2015) han propuesto la convolución como una alternativa a las capas densamente conectadas, con el objetivo de optimizar el

aprendizaje y mejorar su eficiencia.

- ii) **Capa pooling:** Según Voulodimos et al. (2017), la capa pooling tiene la función de reducir la anchura y la altura del volumen de entrada, permitiendo que la siguiente capa convolucional lo procese de manera más eficiente. Sin embargo, la dimensión de profundidad del volumen permanece inalterada. Esta capa lleva a cabo un proceso denominado downsampling o undersampling, mediante el cual se reduce la información, generando una pérdida controlada. A pesar de ello, esta reducción beneficia a la red, ya que disminuye la carga computacional de las capas posteriores y ayuda a prevenir el sobreajuste.
- iii) **La capa fully-connected:** Es la última en un modelo de RNC. En esta capa, cada neurona está completamente conectada con todas las activaciones de la capa anterior, permitiendo la integración de la información procesada. Su activación se obtiene mediante una multiplicación de matrices seguida de una compensación por sesgo.

B. Las redes Neuronales Generativas Adversarias (GAN)

Las *Generative Adversarial Networks* (GAN) fueron propuestas en 2014 por Goodfellow et al. Su principal característica es la combinación de dos redes neuronales que trabajan de manera competitiva: una red generadora, encargada de producir nuevos datos de salida, y una red discriminadora, cuya función es evaluar la calidad de dichas salidas. Esta interacción permite que el modelo mejore progresivamente, generando datos cada vez más realistas.

C. Redes Neuronales Recurrentes (RNN)

Este tipo de red, además de recibir las entradas correspondientes, cuenta con un bucle interno que le permite emplear su propia salida del instante anterior como una nueva entrada (Rumelhart, Hinton & Williams, 1986). Este mecanismo posibilita la persistencia de la información de instantes previos en el modelo, lo que resulta fundamental para el procesamiento de datos

secuenciales, como el reconocimiento de patrones en series temporales o el análisis de texto en lenguaje natural.

A pesar de su gran potencial, las redes neuronales recurrentes (RNN) han presentado desafíos en su entrenamiento y en la modelización de datos con dependencias temporales a largo plazo (Bengio, Simard & Frasconi, 1994). Para mitigar estos problemas, Hochreiter (1997) propuso un nuevo tipo de red recurrente denominada Long Short-Term Memory (LSTM). Este modelo optimiza la velocidad de entrenamiento y mejora la capacidad de representación de dependencias a largo plazo, lo que lo hace especialmente eficaz en aplicaciones como el procesamiento de lenguaje natural y la predicción de series temporales.

2.2.5. Procesamiento del Lenguaje Natural (PLN)

El Procesamiento del Lenguaje Natural (PLN) es un área de la inteligencia artificial que permite a las computadoras comprender, interpretar y generar texto de manera que se asemeje al lenguaje humano (Manning & Schütze, 2018). Su aplicación abarca desde la traducción automática hasta la interacción con asistentes virtuales, lo que ha permitido un avance significativo en la automatización de tareas lingüísticas.

A nivel estructural, el PLN combina modelos de lingüística computacional con técnicas de aprendizaje automático y redes neuronales profundas para procesar grandes volúmenes de texto y extraer información útil (Jurafsky & Martin, 2019). Estas técnicas han sido fundamentales para mejorar la precisión de las aplicaciones de inteligencia artificial en tareas de comunicación escrita y oral.

2.2.5.1. PLN en la clasificación de sentimientos

Una de las aplicaciones más relevantes del PLN en redes sociales es el análisis de sentimientos, que consiste en identificar y clasificar las emociones expresadas en textos digitales (Cambria et al., 2018). Esta técnica permite a las empresas evaluar la percepción del público sobre sus productos o servicios a través de comentarios y publicaciones en plataformas como Twitter y Facebook.

El análisis de sentimientos se basa en métodos supervisados y no supervisados de aprendizaje automático. Los enfoques tradicionales incluyen modelos estadísticos como Naïve Bayes y Support Vector Machines (SVM), mientras que las metodologías más recientes emplean redes neuronales profundas, como Long Short-Term Memory (LSTM) y BERT, que han demostrado mejorar la precisión en la clasificación de sentimientos (Zhang et al., 2018)

2.2.5.2. Detección de patrones y análisis de tendencias en redes sociales

El PLN no solo se limita a la clasificación de sentimientos, sino que también se utiliza para identificar patrones de comportamiento y tendencias en redes sociales. Mediante el análisis de grandes volúmenes de datos textuales, se pueden extraer temas recurrentes y tendencias emergentes en la conversación digital (Qiu et al., 2019).

En este contexto, los modelos basados en Transformers, como BERT y GPT, han demostrado ser altamente efectivos para la extracción automática de información y la identificación de tendencias en tiempo real (Rogers et al., 2021). Esto ha sido clave en sectores como el marketing digital y la comunicación corporativa, donde el análisis de datos de redes sociales permite anticipar cambios en las preferencias de los consumidores.

La detección de patrones en el texto también se aplica en la prevención de noticias falsas y la identificación de contenido potencialmente dañino, contribuyendo a la regulación del ecosistema digital (Oshikawa et al., 2020).

2.2.5.3. Técnicas

- a) **Tokenización:** Técnica fundamental en PLN que permite dividir un texto en sus componentes básicos para su posterior análisis y procesamiento (Otter et al., 2020).
- b) **Lematización:** Es una técnica que consiste en reducir las palabras a su forma canónica o lema. Esta transformación permite normalizar el texto, facilitando su análisis y clasificación posterior (Acosta Guzmán et al., 2024).

- c) **Análisis Sintáctico:** Permite identificar la estructura gramatical de una oración, lo que mejora la comprensión semántica en modelos de PLN (Cambria et al., 2018).

2.2.6. Análisis de sentimientos

El rápido avance tecnológico en las últimas décadas ha impulsado un aumento continuo en la producción de grandes cantidades de datos, los cuales pueden manifestarse en diversos formatos, como texto, imágenes, audio y video. Según estimaciones del sector, aproximadamente el 80% de estos datos son no estructurados, es decir, carecen de una organización definida o de un modelo preestablecido. Además, se calcula que más del 50% de estos datos no estructurados corresponden a información en formato de texto. Este tipo de datos se origina a partir de nuestras interacciones con otras personas y/o sistemas a través de diversas plataformas digitales disponibles en internet, como redes sociales, aplicaciones de mensajería, plataformas de video, comercios electrónicos y asistentes virtuales, entre otros. (Kulkarni & Shivananda, 2019)

2.2.6.1. Modelos de Clasificación en el Análisis de Sentimientos

La clasificación de sentimientos ha evolucionado gracias a la implementación de diversos modelos de aprendizaje automático y aprendizaje profundo, cada uno con características, ventajas y limitaciones particulares.

A. Naive Bayes

Es un modelo probabilístico basado en el teorema de Bayes que asume la independencia condicional entre las características del texto. A pesar de esta suposición simplificadora, se ha demostrado que Naïve Bayes es altamente eficiente en tareas de clasificación de texto, particularmente en dominios donde la rapidez en el procesamiento es prioritaria. Su desempeño puede verse afectado cuando existe alta correlación entre las variables, pero sigue siendo una de las opciones más utilizadas debido a su interpretabilidad y bajo costo computacional (Dey Sarkar et al., 2014). A pesar de su simplicidad, Naïve Bayes sigue siendo un

modelo eficiente para la clasificación de textos debido a su rapidez y bajo costo computacional (Jangra et al., 2020)

B. Support Vector Machines (SVM)

Modelo de clasificación basado en la maximización del margen entre clases dentro de un espacio de alta dimensión. Su capacidad para manejar datos no lineales mediante el uso de funciones kernel lo convierte en una herramienta robusta para el análisis de sentimientos, especialmente en escenarios con grandes volúmenes de datos etiquetados. Estudios recientes han destacado la eficacia de SVM en comparación con otros enfoques tradicionales, mostrando su capacidad para lograr altas tasas de precisión en tareas de clasificación de texto, aunque con tiempos de entrenamiento elevados en conjuntos de datos extensos (Kowsari et al., 2019). Además, los modelos basados en SVM han demostrado ser altamente efectivos en tareas de análisis de sentimientos, debido a su capacidad para manejar datos de alta dimensionalidad y reducir el sobreajuste (Qiu et al., 2019)

C. Bidirectional Encoder Representations from Transformers (BERT)

Ha revolucionado el procesamiento del lenguaje natural al emplear una arquitectura de transformers que captura el contexto de manera bidireccional. A diferencia de los modelos anteriores, BERT no solo considera la información pasada, sino también la futura dentro de una oración, lo que mejora la comprensión semántica. Se ha demostrado que este modelo supera en precisión a las técnicas tradicionales en múltiples tareas de análisis de sentimientos, aunque su principal limitación radica en la alta demanda computacional para su entrenamiento y fine-tuning (Devlin et al., 2019). El autor (Rogers et al., 2021), refuerza la idea anterior cuando menciona que, el modelo BERT ha revolucionado el procesamiento del lenguaje natural al permitir el aprendizaje bidireccional en textos, logrando superar a técnicas tradicionales en múltiples benchmarks de PLN.

D. Long Short-Term Memory (LSTM)

Variante de las redes neuronales recurrentes, están diseñadas para aprender dependencias a largo plazo en secuencias de datos. Gracias a sus mecanismos de puertas de entrada, olvido y salida, las LSTM pueden retener información relevante a lo largo del tiempo, superando los problemas de desvanecimiento de gradiente presentes en las RNN convencionales. Se han aplicado con éxito en la clasificación de sentimientos, mostrando ventajas significativas en la detección de relaciones contextuales y en la mejora de la precisión de los modelos en comparación con enfoques basados en técnicas tradicionales (Yin et al., 2020). Gracias a su capacidad para manejar secuencias de texto y captar dependencias a largo plazo, las LSTM han mejorado significativamente el desempeño en tareas de análisis de sentimientos (Zhang et al., 2018)

2.2.6.2. Métricas en el Análisis de Sentimientos

a) Accuracy

La exactitud del modelo es una de las métricas más utilizadas para evaluar el rendimiento general de los sistemas de clasificación de sentimientos, aunque puede verse afectada en conjuntos de datos desbalanceados (Yadav & Vishwakarma, 2020). Además, indica la proporción de predicciones correctas respecto al total de casos analizados, lo que permite evaluar el rendimiento global del modelo (Wadud et al., 2022).

b) Presición

La precisión mide la cantidad de predicciones correctas dentro de las clasificadas como positivas, siendo una métrica esencial para modelos que deben evitar falsos positivos (Jurafsky & Martin, 2020). En el trabajo de investigación de (Minaee et al., 2021), se enfoca en la capacidad del modelo para identificar correctamente las instancias positivas, minimizando la ocurrencia de falsos positivos y garantizando una clasificación más precisa.

c) Recall

Un alto recall es crucial en tareas donde es más importante detectar correctamente las instancias positivas que minimizar los falsos positivos, como en la detección de comentarios negativos en redes sociales (Zhang et al., 2018). Cabe recalcar que, también mide la habilidad del modelo para recuperar todas las instancias relevantes dentro del conjunto de datos, reduciendo los falsos negativos y mejorando la cobertura de la clasificación (Hassan & Mahmood, 2018).

d) F1-score

Al ser la media armónica entre precisión y recall, el F1-score permite evaluar de manera equilibrada el desempeño del modelo, especialmente en datasets con distribución desigual de clases (Liu & Zhang, 2020). El F1-score por su parte, combina de forma armónica la Precision y el Recall, ofreciendo una visión equilibrada del desempeño del sistema, especialmente en escenarios donde hay un desbalance entre clases.

2.2.7. Sistemas Inteligentes

Según UNED (2017) Los sistemas inteligentes “Se definen como aquellos que presentan un comportamiento externo similar en algún aspecto a la inteligencia humana o animal. Se caracterizan por su capacidad para representar, procesar y modificar de forma explícita conocimiento sobre un problema, y para mejorar su desempeño con la experiencia. Esto les permite resolver problemas concretos determinando las acciones a tomar para alcanzar los objetivos propuestos, a través de la interacción con el entorno y adaptándose a las distintas situaciones”.

2.2.8. Herramientas y Tecnologías para la Analítica de Redes Sociales

2.2.8.1. Bases de datos NoSQL

Las bases de datos NoSQL han surgido como una solución eficaz para gestionar grandes volúmenes de datos no estructurados o semiestructurados, proporcionando escalabilidad y flexibilidad en entornos de alto tráfico como las redes sociales. Estas bases de datos

permiten almacenar y procesar datos de interacciones en tiempo real, lo que las hace esenciales para plataformas como Facebook y Twitter (Khan et al., 2022)

2.2.8.2. PostgreSQL

PostgreSQL es un sistema de gestión de bases de datos relacional orientado a objetos y de código abierto que se destaca por su robustez, extensibilidad y cumplimiento de estándares SQL (Kinsta, 2025)

PostgreSQL ofrece características esenciales como soporte total para transacciones ACID, ejecución concurrente mediante MVCC y compatibilidad con tipos de datos variados (enteros, texto, JSON, binarios), así como extensiones como PostGIS para datos espaciales y PL/pgSQL para lógica procedimental en el servidor (Kinsta, 2025)

2.2.8.3. Lenguajes y bibliotecas para el análisis de datos

A. Python para analítica de redes sociales

Python es uno de los lenguajes más utilizados para la analítica de datos en redes sociales. Python es preferido por su versatilidad y amplia variedad de bibliotecas para minería de datos, mientras que R es reconocido por su enfoque en el análisis estadístico avanzado (García, Luengo & Herrera, 2016, p. 22)

B. TensorFlow y PyTorch

Han revolucionado el aprendizaje profundo al proporcionar herramientas eficientes para la construcción de modelos predictivos, utilizados en la detección de tendencias y análisis de sentimientos en redes sociales (Plà et al., 2022)

C. Scikit-Learn

Es una biblioteca ampliamente utilizada en minería de datos, que ofrece herramientas para clasificación, regresión y clustering en el análisis de datos de redes sociales (Pedregosa et al., 2011, p. 6)

D. NLTK y SpaCy

Son bibliotecas especializadas en Procesamiento del Lenguaje Natural (PLN), utilizadas para analizar y comprender textos generados por usuarios en redes sociales, como comentarios y

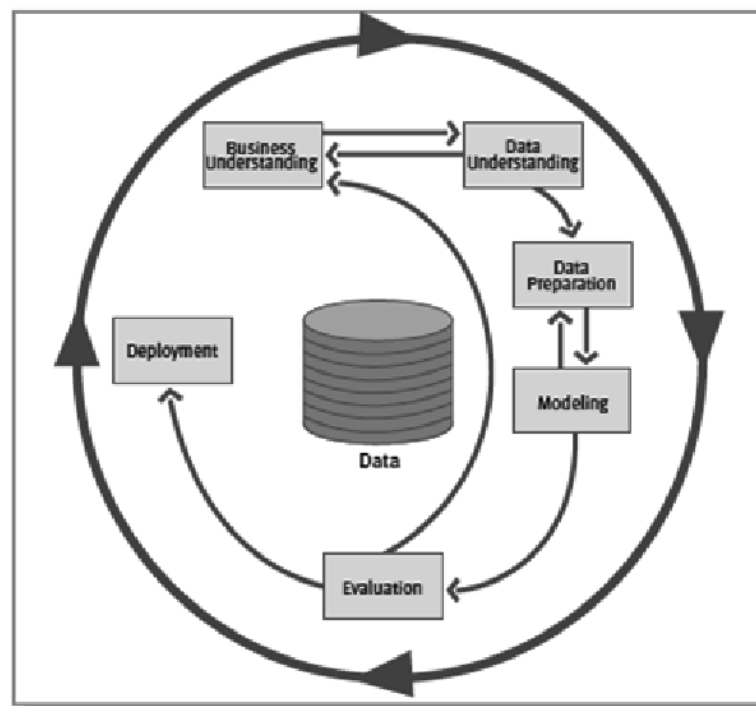
publicaciones (Bird, Klein & Loper, 2009, p. 15; Honnibal & Montani, 2017, p. 3).

2.2.9. Metodología CRISP-DM

El modelo CRISP-DM (Cross Industry Standard Process for Data Mining), propuesto por Chapman año 2000, establece un marco generalizado llevando a cabo proyectos de minería de datos, definiendo sus principales etapas, tareas esenciales y resultados esperados. Asimismo, su guía de usuario proporciona información detallada sobre la implementación del modelo en proyectos específicos, ofreciendo recomendaciones y listas de verificación para cada etapa del proceso. Este método estructura el desarrollo de un proyecto en seis fases interrelacionadas, las cuales se ejecutan de forma iterativa en todo lo relacionado al ciclo de vida del proyecto.

Figura 4:

Ciclo de vida de CRISP DM

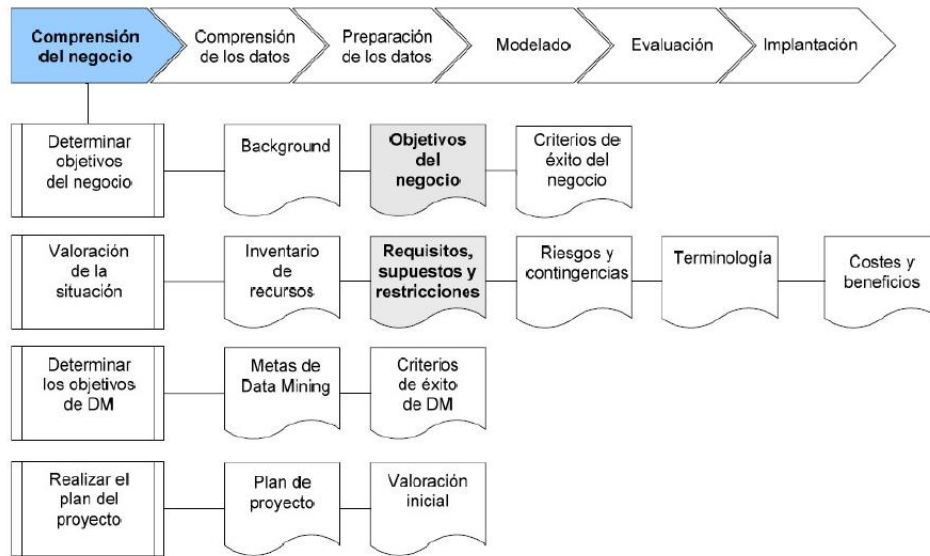


A. Comprensión del negocio

Esta fase inicial es crucial en el proceso, ya que requiere una comprensión profunda de los objetivos y necesidades del proyecto desde una perspectiva empresarial, con el fin de convertirlos en metas técnicas y en un plan de trabajo estructurado.

Figura 5:

Fase de comprensión del negocio

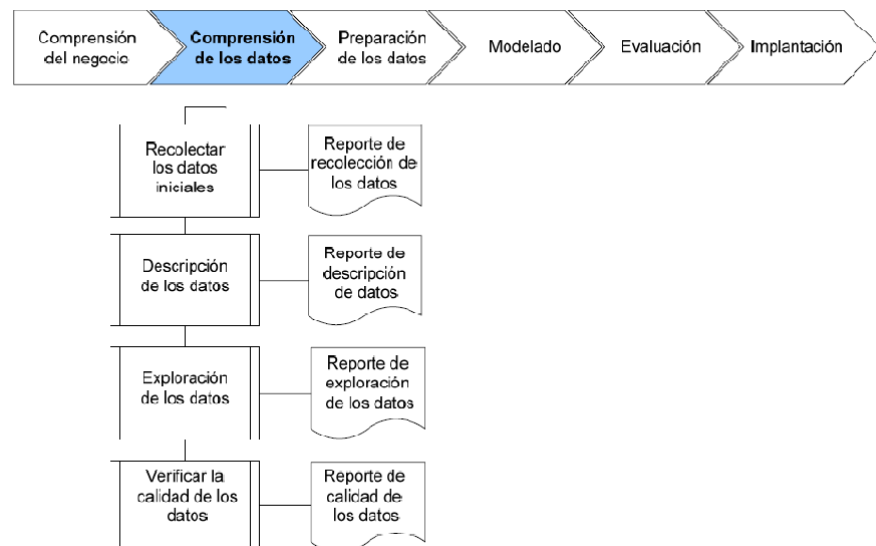


B. Comprensión de los datos

La segunda fase se centra en la recopilación inicial de los datos, permitiendo un primer acercamiento al problema, la familiarización con la información disponible, la evaluación de su calidad y la identificación de relaciones clave para formular las primeras hipótesis.

Figura 6:

Fase de comprensión de los datos



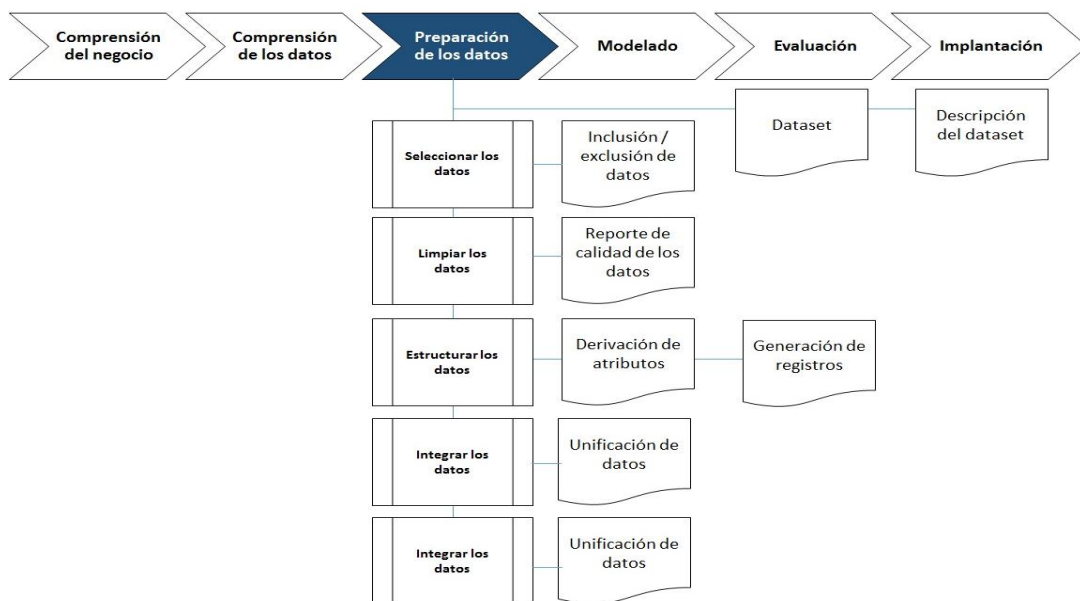
C. Preparación de los datos

Tras la recopilación inicial de los datos obtenidos, se lleva a cabo su preparación para adecuarlos a las técnicas de minería de datos que se aplicarán posteriormente. Estas técnicas pueden incluir la visualización de datos, la identificación de relaciones entre variables o diversas metodologías para su explotación y análisis.

Esta etapa tiene una conexión directa con la fase de modelado, ya que el procesamiento de los datos está determinado por la técnica de modelado seleccionada. Debido a esta interdependencia, las fases de preparación y modelado interactúan continuamente a lo largo del proceso.

Figura 7:

Fase de preparación de los datos

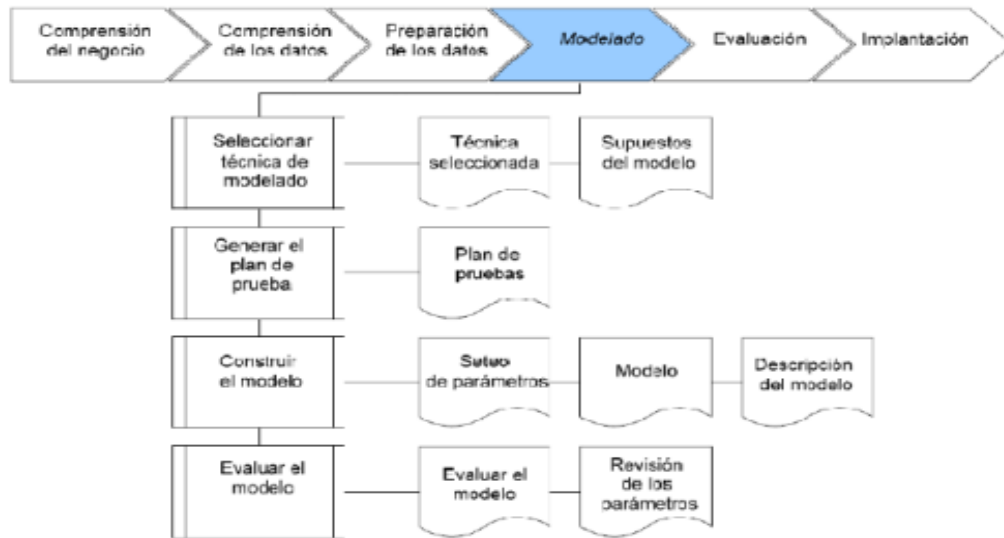


D. Modelado

Durante esta etapa del enfoque CRISP-DM, se seleccionan las estrategias de modelado más adecuadas para el proyecto de minería de datos en cuestión.

Figura 8:

Fase de modelado

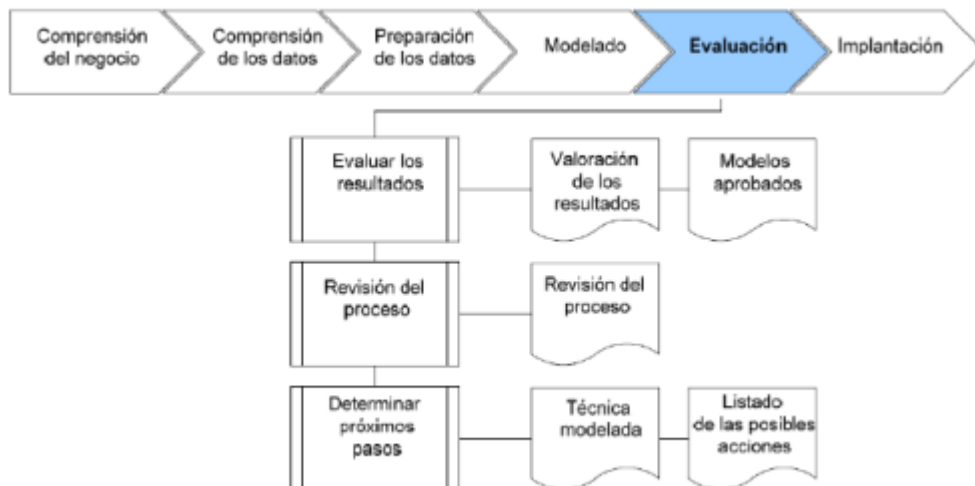


E. Evaluación

En esta fase, se lleva a cabo la evaluación del modelo en función del cumplimiento de los criterios de éxito previamente definidos.

Figura 9:

Fase de evaluación



F. Fase de implantación

Para llevar a cabo la implementación del resultado de la minería de datos dentro de la organización, esta tarea toma en cuenta los resultados obtenidos en la fase de evaluación y define una estrategia de implementación.

2.2.10. Gestión

La gestión dentro de una organización se refiere al conjunto de procesos administrativos y estratégicos que permiten alcanzar objetivos institucionales de manera eficiente (Robbins & Coulter, 2021). Esta implica la planificación, organización, dirección y control de los recursos disponibles para maximizar el rendimiento y la productividad.

2.2.11. Toma de Decisiones

La toma de decisiones es un proceso crítico dentro de la gestión organizacional, ya que implica la selección de la mejor alternativa entre varias opciones disponibles (Simon, 1997). En el ámbito académico, una toma de decisiones eficiente puede impactar positivamente la gestión administrativa, la optimización de recursos y la satisfacción de la comunidad universitaria (Russell & Norvig, 2021).

Según Davenport y Prusak (2018), el uso de analítica avanzada permite reducir el margen de error en las decisiones estratégicas, ya que proporciona información objetiva basada en datos históricos y en modelos predictivos.

III. METODOLOGÍA

3.1. Metodología de investigación

El enfoque de esta investigación es mixto, ya que combina métodos cuantitativos y cualitativos con el objetivo de diseñar e implementar un marco de trabajo para la analítica de redes sociales basado en sistemas inteligentes. Este enfoque es adecuado para investigaciones aplicadas, especialmente en el ámbito de la ingeniería de sistemas e informática, donde es necesario analizar tanto métricas cuantitativas relacionadas con la interacción en redes sociales como interpretar cualitativamente el impacto de estas herramientas en la toma de decisiones estratégicas (Camacho et al., 2021).

Desde la perspectiva cuantitativa, se recopilaron y analizaron grandes volúmenes de datos provenientes de redes sociales institucionales. Este análisis incluyó métricas de interacción, análisis de sentimientos y patrones de comportamiento utilizando herramientas de minería de datos y aprendizaje automático. El uso de técnicas cuantitativas permitió medir con precisión la eficacia del marco de trabajo implementado en términos de eficiencia y mejora en la gestión de datos (Joshi & Kamath, 2023).

Por otro lado, desde la perspectiva cualitativa, se exploraron las percepciones y necesidades de los usuarios del Centro de Cómputo de la Universidad Nacional del Santa, incluyendo personal administrativo, estudiantes y docentes. Estas percepciones se analizaron a través de entrevistas y grupos focales para garantizar que el marco de trabajo propuesto respondiera a las necesidades específicas de la institución. Además, el análisis cualitativo permitió evaluar el impacto del marco en la mejora de la interacción y comunicación con la comunidad universitaria (Viviani et al., 2020).

El enfoque mixto en este caso fue esencial, ya que permitió integrar resultados cuantitativos y cualitativos para desarrollar un marco de trabajo robusto y adaptado al contexto local. Este enfoque no solo proporciona una visión integral del problema, sino que también fortalece la validez de los resultados obtenidos (Creswell & Plano Clark, 2018). La combinación de datos numéricos y contextuales permitió optimizar el diseño y la implementación del marco de trabajo, asegurando su relevancia y aplicabilidad.

3.2. Método de investigación

La investigación "Marco de trabajo para analítica de redes sociales basada en sistemas inteligentes aplicado al Centro de Cómputo de la Universidad Nacional del Santa" es de tipo aplicada debido a que tiene como propósito principal resolver un problema práctico concreto relacionado con la gestión y toma de decisiones estratégicas en el ámbito institucional. Este enfoque busca generar conocimientos que se traduzcan en soluciones tecnológicas, permitiendo implementar sistemas inteligentes para mejorar la analítica de redes sociales. Según Creswell y Creswell (2021), la investigación aplicada se orienta a la resolución de problemas específicos mediante la implementación de conocimientos teóricos y metodológicos en contextos prácticos.

En cuanto a su naturaleza y profundidad, esta investigación se clasifica como descriptiva, explicativa y exploratoria. Es descriptiva porque detalla las características de los datos generados en redes sociales institucionales y los patrones observados mediante su análisis. Además, es explicativa porque busca comprender y determinar las relaciones causales entre la implementación de un marco basado en sistemas inteligentes y la mejora en la gestión y toma de decisiones estratégicas. Finalmente, tiene un componente exploratorio, ya que investiga nuevas formas de integrar técnicas avanzadas de minería de datos y aprendizaje automático en un contexto académico local, donde tales metodologías no han sido implementadas previamente (Camacho et al., 2021).

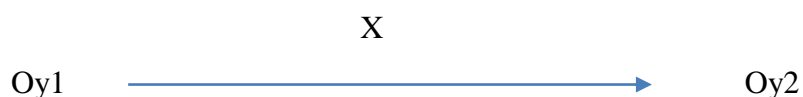
3.3. Diseño de investigación

El diseño de esta investigación se clasifica como cuasi experimental, dado que busca evaluar el impacto de la implementación de un marco de trabajo para analítica de redes sociales basado en sistemas inteligentes en el Centro de Cómputo de la Universidad Nacional del Santa, comparando los resultados antes y después de la intervención. Este subtipo de diseño experimental es apropiado en contextos donde no es posible asignar aleatoriamente a los participantes a un grupo de control o experimental, pero sí se puede observar y medir el efecto de una intervención en condiciones controladas (Creswell & Creswell, 2021).

En este diseño cuasi experimental, se analizaron dos períodos: el primero antes de la implementación del marco de trabajo, donde se recopilaron datos relacionados con la gestión y toma de decisiones estratégicas utilizando métodos tradicionales, y el segundo después de la implementación, donde se evaluaron las mejoras introducidas

por el marco basado en sistemas inteligentes. Este enfoque permitió medir los cambios en variables clave, como el tiempo de respuesta en la gestión de redes sociales, la precisión en la identificación de patrones y la satisfacción de los usuarios internos con las decisiones estratégicas.

La elección del diseño cuasi experimental se fundamenta en que este tipo de investigación aplicada se desarrolla en un entorno institucional donde no es posible ejercer un control total sobre las condiciones experimentales, como lo sería en un laboratorio. Sin embargo, la estructura del diseño cuasi experimental permite establecer relaciones causales entre la intervención y los resultados observados, garantizando la validez interna mediante el uso de herramientas de análisis avanzado (Shadish et al., 2002).



Oy₁ = Observación de la variable dependiente sin marco de trabajo con sistema inteligente

Oy₂ = Observación de la variable dependiente después de la aplicación del marco de trabajo con sistema inteligente

3.4. Población

La población de esta investigación estuvo constituida por los datos generados en las redes sociales oficiales gestionadas por el Centro de Cómputo de la Universidad Nacional del Santa. Estas redes sociales incluyen plataformas como Facebook, Twitter e Instagram, donde se realizan interacciones constantes con estudiantes, docentes, personal administrativo y otros miembros de la comunidad universitaria. Adicionalmente, se consideraron los responsables directos de la gestión y administración de dichas plataformas digitales, quienes están involucrados en los procesos de toma de decisiones estratégicas.

3.5. Muestra

La muestra estuvo compuesta por un dataset con 1000 publicaciones en las redes sociales oficiales y se decidió dividirlo en 50 ejecuciones, cada ejecución consistiría en que el modelo analice un subconjunto de publicaciones de 20 en cada prueba en cada prueba y genere sus respectivas métricas. Además, incluyó al personal del Centro

de Cómputo que interactúa directamente con dichas plataformas y los sistemas de gestión. Se utilizó un muestreo no probabilístico por conveniencia, dado que la selección de los datos y participantes se basó en su accesibilidad y relevancia para el estudio. Este tipo de muestreo fue adecuado debido a las limitaciones de acceso a todos los datos generados y la necesidad de trabajar con un conjunto manejable y representativo de información.

Unidad de análisis:

La unidad de análisis de esta investigación se definió en dos niveles:

- Datos generados en redes sociales: Incluyendo publicaciones, comentarios, reacciones, métricas de interacción y patrones de comportamiento de los usuarios en las plataformas gestionadas por el Centro de Cómputo.
- Personal administrativo del Centro de Cómputo: Específicamente aquellos responsables de la gestión de las redes sociales y la toma de decisiones estratégicas basadas en la información analizada

3.6. Actividades del proceso investigativo

A. Fase de planificación y diagnóstico

- **Revisión bibliográfica:** Se realizó un análisis exhaustivo de literatura científica reciente relacionada con sistemas inteligentes, analítica de redes sociales y marcos de trabajo tecnológicos. Esta actividad permitió identificar teorías, metodologías y herramientas aplicables a la investigación (Camacho et al., 2021; Joshi & Kamath, 2023).
- **Identificación del problema:** Se analizaron las necesidades y limitaciones del Centro de Cómputo en relación con la gestión de sus redes sociales, recopilando información mediante entrevistas con el personal administrativo y el análisis de datos históricos. Definición de objetivos y alcances: Se establecieron objetivos claros y se delimitaron las variables de estudio para garantizar que la investigación abordara las necesidades institucionales.

B. Fase de diseño del marco de trabajo

- **Diseño conceptual:** Se definió la estructura general del marco de trabajo, integrando técnicas de inteligencia artificial (IA), minería de datos y procesamiento del lenguaje natural. Se seleccionaron algoritmos específicos para tareas como análisis de sentimientos, clasificación de datos y detección de patrones.

- **Selección de herramientas tecnológicas:** Se evaluaron y seleccionaron plataformas y tecnologías, como bibliotecas de aprendizaje automático (TensorFlow, PyTorch) y herramientas de análisis de redes sociales (Gephi, NodeXL).

C. Fase de implementación

- **Desarrollo del marco de trabajo:** Se implementó el marco de trabajo propuesto, integrando los sistemas inteligentes seleccionados para procesar los datos obtenidos de las redes sociales institucionales.
- **Pruebas piloto:** Se realizaron pruebas en un entorno controlado para ajustar parámetros, evaluar el rendimiento de los algoritmos y garantizar la compatibilidad con los sistemas existentes del Centro de Cómputo de la UNS.

D. Fase de recolección y análisis de datos

- **Recolección de datos:** Se recopiló información de las redes sociales oficiales durante el período de estudio, incluyendo métricas de interacción, publicaciones y comentarios.
- **Análisis cuantitativo y cualitativo:** Se utilizaron métodos estadísticos y técnicas de minería de datos para evaluar el impacto del marco de trabajo en la gestión y la toma de decisiones estratégicas.

E. Fase de validación y evaluación

- **Validación del marco de trabajo:** Se compararon los resultados obtenidos antes y después de la implementación del marco, utilizando indicadores clave como eficiencia en la gestión, precisión en la identificación de patrones y satisfacción del personal.
- **Análisis de resultados:** Se analizaron los hallazgos para determinar si el marco de trabajo cumplió con los objetivos planteados y para identificar áreas de mejora.

3.7. Técnicas e instrumentos de la investigación

3.7.1. Técnicas de recolección de datos

A. Minería de datos

Esta técnica se utilizó para extraer patrones y tendencias significativas a partir de los datos generados en las redes sociales institucionales. Los algoritmos empleados incluyeron técnicas de clasificación, agrupamiento y análisis de sentimientos.

B. Procesamiento del lenguaje natural (PLN)

Se utilizó para analizar los comentarios, publicaciones y otras formas de texto generado por los usuarios en las redes sociales. El PLN permitió evaluar el sentimiento general de las interacciones, identificar palabras clave y detectar patrones semánticos.

C. Entrevistas semiestructuradas

Se llevaron a cabo con el personal administrativo del Centro de Cómputo de la UNS para comprender las necesidades actuales, identificar las limitaciones de los procesos de gestión y evaluar la percepción del impacto del marco de trabajo implementado.

D. Observación directa

Esta técnica permitió registrar cómo el personal del Centro de Cómputo de la UNS gestionaba las redes sociales antes y después de la implementación del marco de trabajo, lo que ayudó a identificar cambios en la eficiencia y efectividad del proceso.

3.7.2. Instrumentos

- **Minería de datos:** Herramientas de minería de datos y aprendizaje automático como Python (con bibliotecas como Scikit-learn, Pandas y TensorFlow).
- **Procesamiento del lenguaje natural (PLN):** Herramientas como NLTK, SpaCy y el modelo BERT.
- **Entrevistas semiestructuradas:** Guía de entrevista diseñada con base en los objetivos de la investigación.
- **Observación directa:** Lista de verificación estructurada basada en indicadores de gestión.

3.8. Procedimiento para la recolección de datos

A. Definición del alcance y criterios de recolección

- **Selección de plataformas de redes sociales:** Se identificaron las redes sociales oficiales administradas por el Centro de Cómputo de la UNS, incluyendo Facebook, WhatsApp e Instagram, como fuentes principales de datos.

- **Definición del período de análisis:** Se delimitó el período de recolección de datos al intervalo comprendido entre enero y diciembre de 2023, asegurando un conjunto de datos representativo y actualizado.
- **Establecimiento de variables clave:** Se definieron las métricas a analizar, tales como interacciones (likes, comentarios, compartidos), publicaciones, palabras clave y patrones de comportamiento de los usuarios.

B. Preparación de herramientas y sistemas

- **Configuración de herramientas de minería de datos:** Se implementaron y configuraron sistemas basados en Python, empleando bibliotecas como Pandas, TensorFlow y SpaCy para la extracción y procesamiento inicial de los datos.
- **Integración de APIs de redes sociales:** Se estableció conexión con las APIs oficiales de las plataformas seleccionadas para la extracción automatizada de datos, respetando las políticas de privacidad de cada plataforma.
- **Diseño de instrumentos cualitativos:** Se elaboraron guías de entrevistas semiestructuradas y listas de observación para la recopilación de información cualitativa.

C. Recolección de datos automatizada

- **Extracción de datos desde redes sociales:** Se desarrollaron scripts para recolectar publicaciones, comentarios y reacciones de las plataformas seleccionadas. Este proceso incluyó la recopilación de metadatos, como fechas de publicación y volumen de interacciones.
- **Procesamiento inicial de los datos:** Los datos recolectados se almacenaron en bases de datos estructuradas para facilitar su análisis posterior. Se eliminaron registros duplicados y se aplicaron técnicas de limpieza de datos.

D. Aplicación de técnicas cualitativas

- **Entrevistas semiestructuradas:** Se llevaron a cabo entrevistas con los responsables del manejo de las redes sociales en el Centro de Cómputo de la UNS, enfocándose en sus percepciones sobre los procesos actuales y las áreas de mejora.
- **Observación directa:** Se realizó un seguimiento de las actividades diarias relacionadas con la gestión de las redes sociales, utilizando una lista de verificación para registrar las prácticas habituales y los desafíos enfrentados.

E. Validación de la recolección

- **Pruebas piloto:** Antes de la recolección completa, se realizaron pruebas piloto en las redes sociales para verificar la funcionalidad de las herramientas automatizadas y garantizar que los datos recopilados fueran relevantes y completos.
- **Revisión de calidad:** Los datos recolectados se revisaron para identificar errores o inconsistencias, asegurando la calidad y confiabilidad de la información obtenida.

F. Consolidación de los datos

- **Integración de datos cualitativos y cuantitativos:** Los resultados de las entrevistas y observaciones se complementaron con las métricas extraídas de las redes sociales, integrándose en una base de datos única para facilitar el análisis.
- **Almacenamiento seguro:** Toda la información recopilada se almacenó en un sistema de gestión de datos protegido, asegurando la confidencialidad y el acceso restringido.

3.9. Técnicas de procesamiento y análisis de los datos

3.9.1. Procesamiento de datos

A. Limpieza de datos

- Se aplicaron procesos para detectar y eliminar valores nulos, duplicados y outliers en los conjuntos de datos obtenidos de las redes sociales, asegurando la integridad y calidad de la información.
- **Herramientas:** Python (bibliotecas como Pandas y NumPy).

B. Normalización y transformación de datos

- Los datos textuales, como comentarios y publicaciones, se transformaron en un formato estructurado mediante técnicas de procesamiento del lenguaje natural (PLN). Este paso incluyó tokenización, lematización y eliminación de palabras irrelevantes (stop words).
- **Herramientas:** NLTK, SpaCy y el modelo BERT.

C. Estructuración y almacenamiento

- Los datos procesados se almacenaron en una base de datos relacional para facilitar el acceso y análisis. Además, se utilizaron formatos de archivo como CSV y JSON para la interoperabilidad.
- **Herramientas:** MySQL, PostgreSQL.

3.9.2. Análisis de datos

A. Análisis cuantitativo

a) Análisis descriptivo

Se calcularon métricas como frecuencias, promedios, tendencias y tasas de interacción en redes sociales (likes, comentarios, compartidos) para identificar patrones de comportamiento.

Herramientas: Microsoft Excel, SPSS.

b) Análisis de sentimientos

Se clasificaron los comentarios y publicaciones según su polaridad (positivo, negativo, neutro), utilizando modelos de PLN y algoritmos de aprendizaje automático.

Herramientas: TextBlob, TensorFlow, y Scikit-learn.

c) Análisis de correlación

- Se evaluaron las relaciones entre variables clave, como el nivel de interacción en redes sociales y la efectividad en la toma de decisiones estratégicas.
- **Herramientas:** Python (biblioteca Statsmodels).

B. Análisis cualitativo

a) Análisis de contenido

- Se analizaron las respuestas obtenidas en entrevistas semiestructuradas y las observaciones directas, identificando temas recurrentes y percepciones clave relacionadas con la gestión de redes sociales.

b) Codificación temática

- Los datos cualitativos fueron organizados en categorías y temas para facilitar su interpretación y relacionarlos con las métricas cuantitativas.

3.9.3. Validación de resultados

A. Validación cruzada

- Se compararon los resultados obtenidos en diferentes etapas del análisis para asegurar su consistencia y fiabilidad.
- Método: Uso de datos independientes de prueba para validar modelos predictivos.

B. Evaluación del marco de trabajo

- Se realizaron pruebas piloto y análisis comparativo entre los resultados previos y posteriores a la implementación del marco de trabajo. Los indicadores de desempeño fueron evaluados para medir el impacto del sistema inteligente.

3.10. Matriz de Consistencia

Tabla 1:

Matriz de Consistencia

Formulación del Problema	Objetivos	Hipótesis	Variables
Problema general: ¿Cómo puede un marco de trabajo para analítica de redes sociales basado en sistemas inteligentes mejorar la gestión y toma de decisiones en el Centro de Cómputo de la Universidad Nacional del Santa?	Objetivo general: Diseñar e implementar un marco de trabajo para la analítica de redes sociales basado en sistemas inteligentes para mejorar la gestión y la toma de decisiones estratégicas en el Centro de Cómputo de la Universidad Nacional del Santa.	Hipótesis general: La implementación de un marco de trabajo para analítica de redes sociales basado en sistemas inteligentes mejora significativamente la gestión y la toma de decisiones estratégicas en el Centro de Cómputo de la Universidad Nacional del Santa.	Variable independiente: Marco de trabajo para analítica de redes sociales basado en sistemas inteligentes.
Problemas específicos: - ¿Cómo diseñar un marco de trabajo de analítica de redes sociales basado en sistemas inteligentes que optimice la experiencia de los usuarios y docentes en el centro de cómputo de la Universidad Nacional del Santa? - ¿Qué factores tecnológicos limitan la eficiencia de las redes sociales del centro de cómputo de la Universidad Nacional del Santa, y cómo pueden ser mejorados para optimizar su rendimiento? - ¿Cuáles son los aspectos tecnológicos clave que afectan la eficacia de las redes sociales del centro de cómputo de la Universidad Nacional del Santa, y cómo se pueden optimizar para lograr una mayor efectividad en su uso? - ¿Cuáles son los factores que afectan la calidad del análisis de datos de las redes sociales del centro de cómputo de la Universidad Nacional del Santa, y qué mejoras tecnológicas o metodológicas pueden implementarse para mejorar la precisión y utilidad de estos análisis?	Objetivos específicos: - Proponer y Diseñar un marco de trabajo de analítica de redes sociales basado en sistemas inteligentes orientado a la experiencia de los usuarios y docentes en el centro de cómputo de la Universidad Nacional del Santa. - Aumentar la eficiencia tecnológica de las redes sociales del centro de cómputo de la Universidad Nacional del Santa. - Elevar la eficacia tecnológica de las redes sociales del centro de cómputo de la Universidad Nacional del Santa. - Incrementar la calidad del análisis de datos de las redes sociales del centro de cómputo de la Universidad Nacional del Santa. - Acrecentar la detección de patrones y tendencias de las redes sociales del centro de cómputo de la Universidad Nacional del Santa.	Hipótesis específicas: - La implementación de un marco de trabajo de analítica de redes sociales basado en sistemas inteligentes mejora la experiencia de los usuarios y docentes en el centro de cómputo de la Universidad Nacional del Santa. - La mejora de los componentes tecnológicos, como la infraestructura, las herramientas de análisis de datos y la integración de sistemas inteligentes, aumenta la eficiencia de las redes sociales del centro de cómputo de la Universidad Nacional del Santa. - La optimización de los procesos tecnológicos incrementa la eficacia de las redes sociales del centro de cómputo de la Universidad Nacional del Santa. - La implementación de herramientas avanzadas de procesamiento de datos, algoritmos de inteligencia artificial y metodologías de análisis más robustas mejora significativamente la calidad del análisis de datos de las redes sociales del centro de cómputo de la Universidad Nacional del Santa.	Variable dependiente: Gestión y toma de decisiones estratégicas en el Centro de Cómputo de la Universidad Nacional del Santa.

3.11. Operacionalización de las Variables

Tabla 1:

Operacionalización de las Variables

Variable	Definición Conceptual	Dimensión	Indicadores	Instrumento	Medida
Variable Independiente: Marco de trabajo para analítica de redes sociales basado en sistemas inteligentes	Sistema estructurado que utiliza tecnologías de inteligencia artificial y técnicas avanzadas para analizar información de redes sociales.	Diseño del marco	Propuesta de Metodología para Marco de trabajo con analítica de redes sociales	Checklist de documentación, Evaluación técnica de documentos	Documentos
		Calidad del Análisis de Datos	Comparación de Precisión en Modelos de Análisis de Sentimientos	Métricas del sistema, Reportes de pruebas piloto	Porcentaje (%)
		Eficiencia tecnológica	Tiempo promedio de procesamiento	Registro de tiempos del sistema	Segundos
		Eficacia Tecnológica	Precisión en los análisis	Registros de decisiones, Encuestas de precisión y utilidad	Porcentaje (%)
Variable Dependiente: Gestión y toma de decisiones estratégicas en el Centro de Cómputo de la Universidad Nacional del Santa	Proceso mediante el cual se administran recursos, operaciones y datos para optimizar el desempeño institucional y apoyar decisiones estratégicas.	Dimensión Patrones y Tendencias	Cantidad de Insights	Registro del sistema, Análisis cualitativo	Cantidad (Número de Insights)

IV. RESULTADOS Y DISCUSIÓN

4.1. Resultados

4.1.1. Metodología Propuesta para el Marco de Trabajo de Analítica de Redes Sociales con Sistemas Inteligentes

La metodología CRISP-SNA es una adaptación del modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), pero con un enfoque específico en el análisis de redes sociales (SNA, por sus siglas en inglés Social Network Analysis) con el uso de sistemas inteligentes. Este modelo ayuda a abordar proyectos de análisis de redes sociales de manera estructurada, proporcionando un marco para extraer conocimiento de los datos relacionados con las interacciones, relaciones y conexiones entre entidades (personas, organizaciones, etc.).

La metodología CRISP-SNA propuesta sigue una serie de pasos que se aplican específicamente a la analítica de redes sociales, lo cual facilita el análisis de las relaciones y comportamientos dentro de una red. A continuación, se detalla las fases que componen el proceso (Ver Anexo 05):

4.1.1.1. Fase 1: Comprensión del Dominio y Definición del Problema (Business Understanding)

En esta fase se define el problema a resolver, se establecen los objetivos y se identifican los requerimientos del análisis de redes sociales con sistemas inteligentes mediante procesamiento de lenguaje natural (PLN) y análisis de sentimientos en Facebook e Instagram.

A. Propósito del Análisis

- Propósito General

El objetivo principal es desarrollar un marco de trabajo para la analítica de redes sociales que sea aplicable específicamente al Centro de Cómputo de la Universidad Nacional del Santa (UNS), utilizando sistemas inteligentes basados en procesamiento de lenguaje natural (PLN). Este marco debe permitir la evaluación y análisis de las interacciones y

opiniones generadas en las redes sociales en relación con la UNS, con el fin de optimizar la toma de decisiones estratégicas para mejorar los servicios y la gestión del Centro de Cómputo.

- **Propósitos específicos:**

a) Desarrollar un marco de trabajo basado en PLN y análisis de sentimientos

Crear una arquitectura flexible y escalable que permita analizar datos de redes sociales con técnicas de PLN, con el objetivo de identificar sentimientos y opiniones sobre el Centro de Cómputo de la UNS.

b) Aplicar técnicas de análisis de redes sociales (SNA)

Evaluar cómo las interacciones en redes sociales (como menciones, comentarios, hashtags) se distribuyen entre los diferentes actores involucrados con la UNS (estudiantes, docentes, personal administrativo).

c) Optimizar la gestión del Centro de Cómputo

Analizar el impacto de los comentarios, opiniones y sentimientos de los usuarios (estudiantes, docentes) en relación con los servicios proporcionados por el Centro de Cómputo

d) Proveer herramientas de soporte para la toma de decisiones

Desarrollar informes y dashboards visuales que permitan a los administradores y responsables del Centro de Cómputo tomar decisiones basadas en datos concretos de redes sociales.

B. Identificación de Stakeholders (interesados)

- **Director del Centro de Cómputo de la UNS:**

Necesita comprender cómo la comunidad universitaria percibe sus servicios, para poder mejorar la infraestructura, la oferta de servicios y la atención al usuario.

- **Estudiantes y docentes de la UNS:**

Son los usuarios directos de los servicios del Centro de Cómputo. Sus opiniones, comentarios y sentimientos en redes sociales servirán como base para el análisis.

- **Investigadores y académicos en ingeniería de sistemas e informática:**

Pueden utilizar este marco de trabajo como una referencia para futuras investigaciones en el ámbito de la analítica de redes sociales, PLN y sistemas inteligentes.

C. Importancia

- **Académica y Tecnológica**

- La analítica de redes sociales es una herramienta poderosa para entender la opinión y el comportamiento de los usuarios.
- El uso de PLN y análisis de sentimientos permite obtener información cualitativa valiosa sobre la percepción de los servicios del Centro de Cómputo.
- La combinación de SNA + PLN facilita la identificación de problemas, tendencias y oportunidades de mejora

- **Impacto en la Gestión del Centro de Cómputo**

- Ayudará a los administradores del Centro de Cómputo a mejorar su comunicación digital y optimizar sus servicios.
- Permitirá identificar problemas recurrentes en los servicios a partir de los comentarios negativos.
- Facilitará la toma de decisiones estratégicas mediante la visualización de tendencias y métricas clave.

D. Alcance

- **Universidad:** Universidad Nacional del Santa.
- **Entidad Analizada:** Centro de Cómputo de la Universidad Nacional del Santa.
- **Redes Sociales:** Facebook e Instagram.

- **Población Analizada:** Estudiantes, docentes y administrativos que interactúan en redes sociales sobre temas del Centro de Cómputo.

E. Plataformas de Redes Sociales a Estudiar

El CECOMP ha utilizado varias redes sociales para su estrategia de marketing digital:

- **Facebook:** Red principal con 8188 seguidores y la mayor interacción.
- **Instagram:** En crecimiento, pero con menor impacto en la conversión de alumnos.
- **WhatsApp:** Usado para comunicación directa con alumnos.

Foco del estudio: Facebook e Instagram, ya que son las plataformas con mayor impacto en la interacción y conversión de alumnos.

F. Indicadores Clave de Rendimiento (KPIs)

Estos KPIs permitirán medir el impacto del marco de trabajo basado en sistemas inteligentes.

Tabla 2:

Indicadores Claves de Rendimiento

KPI	Descripción	Métrica
Distribución de sentimientos	de Clasificación de comentarios en positivo, negativo o neutro.	% por categoría.
Tendencia de Sentimientos	Variación del sentimiento predominante en el tiempo.	Cambio porcentual semanal/mensual.
Engagement Sentimental	Relación entre sentimientos y nivel de interacción.	% de engagement por categoría de sentimiento
Conversión basada en sentimientos	Relación entre comentarios positivos y número de inscripciones.	% de conversión de seguidores a estudiantes.
Tiempo de respuesta optimizado	Reducción del tiempo de respuesta basado en detección automática de sentimientos.	Minutos por respuesta

G. Contexto del Problema

El Centro de Cómputo de la Universidad Nacional del Santa, como muchas otras instituciones educativas, enfrenta desafíos en la gestión de la información proveniente de redes sociales sobre sus servicios, infraestructura y actividades, tales como:

- Dificultad para interpretar la percepción de los usuarios, ya que no hay un análisis estructurado de los comentarios.
- Falta de una estrategia basada en datos para mejorar la comunicación con la comunidad universitaria.
- Ausencia de herramientas automatizadas para responder a consultas y retroalimentación en tiempo real.
- Desconocimiento sobre el impacto de los sentimientos en la conversión de seguidores en estudiantes.

El análisis de sentimientos mediante PLN permitirá extraer conocimiento útil a partir de los comentarios en redes sociales y mejorar las estrategias de captación y fidelización de alumnos.

H. Preguntas clave para orientar el análisis

- ¿Qué emociones predominan en los comentarios de las redes sociales del CECOMP?
- ¿Cómo afectan los sentimientos positivos o negativos a la inscripción de estudiantes?
- ¿Cómo varía la percepción del CECOMP en función de las campañas publicitarias en redes sociales?
- ¿Es posible automatizar respuestas basadas en el análisis de sentimientos para mejorar la interacción con los usuarios?

I. Metodología y Enfoque

El enfoque del análisis se centrará en el uso de procesamiento de lenguaje natural (PLN) para extraer sentimientos y emociones de los comentarios y publicaciones en redes sociales. Se utilizarán herramientas de análisis de sentimientos como VADER, TextBlob, o modelos preentrenados en Deep Learning como

BERT para clasificar los sentimientos expresados. Posteriormente, se empleará análisis de redes sociales (SNA) para mapear las interacciones y relaciones entre los usuarios, utilizando herramientas como Gephi o NetworkX en Python.

J. Restricciones y Limitaciones

- Acceso a datos: Facebook e Instagram tienen restricciones en sus APIs, lo que puede limitar la cantidad de datos disponibles.
- Idioma y contexto: Los modelos de PLN deben ser ajustados para entender lenguaje informal y regionalismos.
- Sesgo en los datos: Algunos grupos de usuarios pueden estar sobrerrepresentados en el análisis.
- Ironía y sarcasmo: Puede dificultar la clasificación precisa del sentimiento.

K. Actores claves y responsabilidades

Tabla 3:

Actores claves y responsabilidades

Actor	Rol	Responsabilidad
Investigador	Responsable del proyecto	Definir la metodología y desarrollar el sistema inteligente.
Centro de Cómputo UNS	Entidad analizada	Proporcionar información sobre servicios y comunicación digital.
Usuarios de redes sociales	Fuente de datos	Interactúan y generan opiniones en Facebook e Instagram.
Asesor	Supervisor	Brinda retroalimentación y guía en el desarrollo del estudio.

4.1.1.2. Fase 2: Comprensión de los datos de Redes Sociales (Data Understanding)

La Fase 2 es crucial, ya que establece las bases para el análisis adecuado de la información proveniente de redes sociales. En el contexto de la tesis doctoral titulada “Marco de trabajo para analítica de redes sociales basada en sistemas inteligentes aplicado al Centro de Cómputo de la Universidad Nacional del Santa”, con el uso de procesamiento de lenguaje natural (PLN) y análisis de sentimientos, esta fase se enfocará en la recolección, exploración y evaluación de los datos disponibles, que provienen principalmente de interacciones en redes sociales relacionadas con el Centro de Cómputo de la UNS

A. Fuentes de Datos

Para realizar el análisis de sentimientos en redes sociales del Centro de Cómputo de la Universidad Nacional del Santa (CECOMP), se identificarán las principales fuentes de datos:

- **Plataformas de redes sociales a analizar:**
 - ✓ Facebook: Principal canal de comunicación con más de 8188 seguidores.
 - ✓ Instagram: Plataforma de interacción visual con creciente participación.
- **Tipos de datos a recolectar:**
 - ✓ Publicaciones oficiales: Contenido generado por el CECOMP (promociones, anuncios, eventos).
 - ✓ Comentarios de los usuarios: Opiniones, preguntas y feedback en redes sociales.
 - ✓ Reacciones e interacciones: Likes, shares, emojis y engagement en cada publicación.
 - ✓ Mensajes directos: Consultas y respuestas de usuarios en redes sociales.
- **Herramientas para la extracción de datos:**
 - ✓ Facebook Graph API: Para obtener publicaciones, comentarios y reacciones.
 - ✓ Instagram API: Para recopilar interacciones y comentarios en posts.

- ✓ Web Scraping (si es necesario): Utilización de Selenium o BeautifulSoup en Python.

Nota: Se garantizará el cumplimiento de políticas de privacidad y términos de uso de cada plataforma.

B. Características de los Datos

▪ Tipos de datos y su formato

Tabla 4:

Tipos de datos y formato

Tipo de Dato	Ejemplo	Formato	Descripción
Texto	“Este curso es excelente, me encantó”	JSON, CSV	Contenido de publicaciones y comentarios
Emojis	😊 😞 🎓	Unicode	Caracteres especiales
Número de interacciones	500 likes, 100 comentarios	Entero	Número de me gusta, compartidos, menciones
Fecha y Hora	2025-02-06 14:30:00	Timestamp	Momento en que se publicó el contenido
Usuario	@JuanPerez	String (URL)	Perfil del autor del comentario (anonimizado)
URL de publicación	https://facebook.com/photo/?fbid=8598614692247&set=a.128275	String (URL)	Clasificación automática del texto (positivo, negativo, neutro)
Sentimiento	Positivo	String	

▪ Aspectos a considerar:

- Los comentarios pueden contener errores ortográficos, sarcasmo y abreviaciones, lo que requiere preprocesamiento de texto.

- Se analizará la frecuencia de interacciones para evaluar el impacto de cada publicación.
- Los emojis serán considerados en el análisis de sentimientos, ya que pueden representar emociones.

C. Recolección de datos

Una vez identificadas las fuentes de datos, se debe proceder a la recolección de los mismos. Este proceso incluye:

- **Extracción de datos de redes sociales:**

- Usar APIs oficiales de redes sociales como Instagram API, Facebook Graph API o herramientas como Scrapy para obtener datos públicos de las plataformas. Los datos de interés incluyen: menciones, hashtags (#CentroDeComputoUNS, #CecompUNS, #Cecomp), comentarios en publicaciones, likes, retweets, etc.
- En caso de utilizar herramientas como BeautifulSoup o Selenium, se pueden extraer datos web de otras plataformas o foros.

- **Recopilación de datos de retroalimentación interna:**

Si existen encuestas previas, se puede acceder a los resultados obtenidos por el Centro de Cómputo de la UNS o las encuestas de satisfacción de estudiantes y docentes.

- **Almacenamiento de datos:**

Los datos obtenidos deben ser almacenados en una base de datos adecuada o en archivos estructurados (como CSV, JSON, SQL) que permitan un fácil procesamiento posterior.

D. Evaluación de la Calidad de los Datos

Para garantizar calidad de datos, se evaluarán los aspectos:

- ✓ **Consistencia:** Se verificará que los datos no contengan valores atípicos o erróneos.
- ✓ **Compleitud:** Se analizará si hay comentarios sin texto o

publicaciones con metadatos faltantes.

- ✓ **Relevancia:** Se excluirán comentarios irrelevantes (spam, publicidad externa).
- ✓ **Duplicidad:** Se eliminarán comentarios repetidos o respuestas idénticas.
- ✓ **Sesgo en los datos:** Se analizará si ciertos grupos tienen mayor representación en los comentarios.
- ✓ **Técnicas utilizadas:**
 - Eliminación de duplicados usando `drop_duplicates()` en Pandas.
 - Corrección de caracteres especiales y ortografía con TextBlob.
 - Eliminación de spam con listas negras de palabras clave.

Tabla 5:

Problemas Potenciales en los Datos y Soluciones

Problema	Descripción	Solución
Datos incompletos	Publicaciones sin suficientes interacciones o contexto.	Filtrar publicaciones con menos de un número mínimo de interacciones.
Spam y ruido	Comentarios irrelevantes o repetitivos.	Eliminación de publicaciones automáticas o irrelevantes.
Ironía y sarcasmo	Puede afectar la clasificación del sentimiento.	Uso de modelos avanzados de PLN como BERT o RoBERTa.
Sesgo en los datos	Algunos grupos pueden estar sobre-representados.	Normalizar la muestra para representar diferentes tipos de usuarios.

E. Exploración de los Datos (EDA - Exploratory Data Analysis)

Antes de aplicar el análisis de sentimientos, se realizará un análisis exploratorio para comprender la distribución de los datos.

■ Técnicas de análisis exploratorio:

- 1) Frecuencia de palabras en los comentarios mediante nubes de palabras (word clouds).
- 2) Distribución de sentimientos preliminar (polaridad positiva, negativa, neutra).

- 3) Análisis de emojis para identificar patrones de emociones en los usuarios.
- 4) Análisis de la actividad por día/hora, para encontrar los momentos de mayor interacción.
- 5) Mapas de calor de engagement, que muestran qué tipo de publicaciones generan más interacción.

▪ **Herramientas utilizadas:**

- NLTK y SpaCy para el análisis léxico.
- Matplotlib y Seaborn para visualización de datos.
- Pandas para estadísticas descriptivas.

Figura 10:

Código Python para gráfico de distribución de sentimientos

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Cargar datos desde el archivo CSV
6 df = pd.read_csv("comentarios_cecomp.csv")
7
8 # Verificar las primeras filas para revisar la estructura del archivo
9 print(df.head())
10
11 # Suponiendo que la columna de sentimientos se llama "Sentimiento"
12 # Si el nombre es diferente, ajústalo según el archivo CSV
13 conteo_sentimientos = df["Sentimiento"].value_counts()
14
15 # Configurar el estilo del gráfico
16 sns.set_style("whitegrid")
17 plt.figure(figsize=(8, 6))
18
19 # Crear un gráfico de barras
20 ax = sns.barplot(x=conteo_sentimientos.index, y=conteo_sentimientos.values,
21                 palette={"Positivo": "green", "Negativo": "red", "Neutro": "gray"})
22
23 # Agregar etiquetas y título
24 plt.xlabel("Sentimiento", fontsize=12)
25 plt.ylabel("Frecuencia", fontsize=12)
26 plt.title("Distribución de Sentimientos en Redes Sociales", fontsize=14)
27
28 # Mostrar valores sobre las barras
29 for p in ax.patches:
30     ax.annotate(f"{p.get_height()}", (p.get_x() + p.get_width() / 2, p.get_height()),
31               ha='center', va='bottom', fontsize=12)
32
33 # Mostrar el gráfico
34 plt.show()
35

```

F. Preparación de los datos

Una vez que los datos se han evaluado, se deben preparar para el análisis mediante técnicas de procesamiento de lenguaje natural (PLN) y análisis de redes sociales. Esta fase de preparación implica varias acciones clave:

▪ **Limpieza de datos:**

- Eliminación de duplicados: Si hay registros duplicados, es necesario eliminarlos.
- Eliminación de ruido: Filtrar los datos irrelevantes, como menciones o publicaciones que no están relacionadas con el Centro de Cómputo o que son demasiado genéricas.
- Corrección de errores ortográficos: En textos que provienen de redes sociales, es común encontrar errores ortográficos o abreviaturas. Utilizar algoritmos de corrección de texto o transformaciones de normalización

▪ **Transformación de los datos:**

- Tokenización: El primer paso en el procesamiento de texto es la tokenización, que consiste en dividir el texto en unidades más pequeñas (tokens) como palabras o frases clave.
- Eliminación de palabras vacías (stop words): Las palabras comunes que no aportan valor semántico (como “el”, “de”, “en”) deben ser eliminadas.
- Lematización o stemming: Es útil reducir las palabras a su raíz (por ejemplo, "mejorar" y "mejorando" se reducirían a “mejor”).

▪ **Análisis de sentimientos:**

- Clasificación de sentimientos: Utilizar modelos de PLN para clasificar los sentimientos de los textos como positivos, negativos o neutrales. Herramientas como VADER, TextBlob, o modelos más avanzados como BERT pueden ser útiles en este paso.

- **Creación de matrices de relaciones**

- Red de interacciones: Si el análisis incluye relaciones entre los usuarios (por ejemplo, menciones o retweets), se pueden construir matrices de adyacencia o grafos que representen la red de interacciones entre los actores (usuarios, entidades).

G. Almacenamiento de los Datos

Los datos extraídos serán almacenados en una base de datos adecuada para su análisis posterior.

- **Opciones de almacenamiento:**

- Base de datos relacional (MySQL, PostgreSQL): Para almacenar los comentarios estructurados.
- Base de datos NoSQL (MongoDB): Para manejar datos no estructurados con texto y emojis.
- Data Lake en Google Cloud Storage: Para grandes volúmenes de datos históricos.

- **Esquema de la base de datos:**

Tabla 6:

Almacenamiento de datos

Campo	Tipo de Dato	Descripción
id_comentario	INT (PK)	Identificador único
texto	TEXT	Contenido del comentario
sentimiento	VARCHAR(10)	Positivo, Negativo, Neutro
usuario	VARCHAR(50)	Nombre de usuario
fecha	TIMESTAMP	Fecha de publicación
plataforma	VARCHAR(20)	Facebook o Instagram
interacciones	INT	Número de likes/reacciones

H. Evaluación de la representatividad de los datos

En esta etapa, se analiza si los datos recopilados de Facebook e Instagram reflejan de manera adecuada la percepción y las interacciones de la comunidad universitaria en relación con los servicios del Centro de Cómputo de la Universidad Nacional del Santa. Esto es fundamental para garantizar que las conclusiones obtenidas sean válidas y aplicables a toda la comunidad universitaria, evitando sesgos que puedan distorsionar el análisis de sentimientos y las redes de interacción. Algunas preguntas clave son:

- ¿Los datos recopilados son representativos del comportamiento de toda la comunidad universitaria? Se debe considerar si los datos cubren una amplia gama de usuarios (estudiantes, docentes, administrativos) y si las interacciones cubren todos los servicios y aspectos relevantes del Centro de Cómputo.
- ¿Existen sesgos en los datos? Es importante identificar si algunos grupos de usuarios (por ejemplo, los estudiantes más activos en redes sociales) están sobrerrepresentados en comparación con otros.

1) Representatividad de los Datos en los estudiantes

Para evaluar si los datos son representativos, se deben considerar los siguientes aspectos:

a. Diversidad de Usuarios en la Muestra

Es crucial que los datos reflejen la participación de todos los actores dentro de la universidad. Los principales grupos de interés son:

- **Estudiantes:** Son los principales usuarios de los servicios del Centro de Cómputo. Se analizará si los comentarios provienen de estudiantes de diferentes facultades y niveles académicos.

- **Docentes:** Utilizan los servicios tecnológicos para actividades académicas. Se verificará si sus interacciones están bien representadas en los datos.
- **Administrativos:** Personal que gestiona el uso de los recursos del Centro de Cómputo de la UNS. Se evaluará si su opinión sobre la gestión tecnológica y académica está incluida.
- **Otros Usuarios:** Existen interacciones de exalumnos o colaboradores externos que también pueden aportar información valiosa.
- ✓ **Medida de control:** Se generará un perfil de usuario a partir de la descripción de los perfiles públicos (cuando sea posible) y se analizará la distribución de usuarios por categoría.

b. Cobertura de Servicios y Temas Relevantes

El Centro de Cómputo ofrece diversos servicios, por lo que se debe evaluar si los comentarios y publicaciones cubren todos los aspectos importantes:

- Acceso al aula virtual.
- Horarios y Matrículas.
- Soporte técnico y atención al usuario.
- Cursos Libres y de Especialización.
- Certificados.
- ✓ **Medida de control:** Se aplicará clasificación de temas (Topic Modeling) usando LDA (Latent Dirichlet Allocation) para verificar qué temas predominan en los datos y si existe algún servicio con baja representación en la muestra.

2) Identificación de Sesgos en los Datos

a. Sesgo por Sobrerrepresentación de Usuarios Activos

En redes sociales, un pequeño grupo de usuarios tiende a generar la mayoría del contenido, lo que puede distorsionar el análisis. Para identificar si esto ocurre:

- Se calculará la distribución de actividad de los usuarios (número de publicaciones/comentarios por usuario).
- Se verificará si una minoría de usuarios representa la mayoría de las interacciones, lo que podría sesgar la opinión general.
- ✓ **Medida de control:** Se aplicará una técnica de normalización de datos, ponderando las opiniones de usuarios menos activos para evitar que una minoría defina el análisis general.

b. Sesgo Temporal

Las interacciones en redes sociales Facebook e Instagram pueden estar influenciadas por eventos específicos (problemas técnicos, cortes de internet, anuncios importantes), lo que puede dar una visión distorsionada si solo se analizan ciertos periodos.

- ✓ **Medida de control:**
 - Se verificará si los datos cubren un período representativo (mínimo 6 meses a 1 año).
 - Se utilizará análisis de series temporales para evaluar si las opiniones varían en función del tiempo o eventos específicos.

c. Sesgo de Opiniones Extremas

En redes sociales, los usuarios tienden a compartir más opiniones cuando tienen experiencias muy positivas o muy negativas, lo que puede afectar la evaluación global del sentimiento.

- ✓ **Medida de control**
 - Se comparará la distribución de comentarios positivos, negativos y neutros para verificar si hay un predominio de opiniones extremas.
 - Se ajustarán los modelos de análisis de sentimientos para mejorar la detección de opiniones neutras.

d. Sesgo en la Interpretación del Lenguaje

El análisis de sentimientos mediante PLN puede verse afectado por expresiones informales, ironía y regionalismos, lo que podría generar errores en la clasificación.

✓ Medida de control:

- Se entrenará el modelo de análisis de sentimientos con un corpus adaptado a la realidad lingüística de los estudiantes de la universidad.
- Se implementará un módulo de detección de sarcasmo para reducir errores en la clasificación de sentimientos.

3) Estrategias para Mejorar la Representatividad

Si se detecta que los datos no son representativos, se aplicarán las siguientes estrategias:

Tabla 7:

Estrategias para mejorar la representatividad

Problema Detectado	Solución Propuesta
Falta de participación de docentes y administrativos	Ampliar la recopilación de datos en grupos de docentes y páginas institucionales.
Algunos servicios del Centro de Cómputo tienen poca representación	Filtrar datos por palabras clave y buscar menciones indirectas en redes.
Predominio de usuarios muy activos	Normalizar las opiniones ponderando la participación.
Sesgo temporal (eventos negativos recientes)	Ampliar el período de análisis y promediar resultados en el tiempo.

4.1.1.3. Fase 3: Preparación de los datos para el Análisis (Data Preparation)

La preparación de datos es una de las etapas más críticas en el análisis de redes sociales, ya que garantiza que los datos sean consistentes, limpios y estructurados para su procesamiento mediante algoritmos de Procesamiento de Lenguaje Natural (PLN) y análisis de sentimientos.

1) Selección de Datos Relevantes

Se filtran y seleccionan los datos recopilados en la fase anterior para enfocarse en información clave relacionada con el Centro de Cómputo de la UNS.

Los Criterios de selección son:

- Publicaciones y comentarios que mencionen palabras clave relacionadas con los servicios del Centro de Cómputo.
- Interacciones directas entre usuarios que discutan temas relevantes.
- Datos con información suficiente para el análisis de sentimientos (texto significativo, no solo emojis o frases cortas).
- Mensajes con metadatos completos, incluyendo fecha, usuario (anonimizado) y número de interacciones.
- **Medida de control:** Uso de filtrado basado en palabras clave y modelos de clasificación de texto.

2) Limpieza de Datos

- **Eliminar duplicados:** Identificación y remoción de comentarios repetidos.
- **Manejo de valores nulos:** Se eliminarán comentarios sin texto o con contenido vacío.
- **Eliminación de caracteres especiales:** Se removerán símbolos innecesarios como #, @, &, a excepción de hashtags relevantes.
- **Corrección de codificación de caracteres:** Evitar errores en caracteres especiales como ñ o acentos.
- **Filtrado de spam:** Identificación de contenido promocional o irrelevante mediante listas negras de palabras.

Tabla 8:*Tareas de limpieza*

Problema	Acción Correctiva	Herramientas
Datos duplicados	Eliminar mensajes repetidos o compartidos varias veces.	Pandas
Mensajes muy cortos o sin contexto	Descartar textos con menos de 5 palabras útiles.	NLTK, SpaCy
Caracteres especiales y URLs	Remover enlaces, emojis y símbolos innecesarios.	Expresiones Regulares (Regex)
Lenguaje informal, errores ortográficos	Normalizar ortografía y corregir palabras mal escritas.	SymSpell, Hunspell
Spam o publicidad	Identificar y filtrar mensajes no relevantes.	Modelos de detección de spam

Figura 11:*Código Python para limpieza de datos*

```

1 import pandas as pd
2 import re
3 import nltk
4 from nltk.corpus import stopwords
5 from nltk.tokenize import word_tokenize
6
7 # Descargar recursos necesarios de NLTK
8 nltk.download('stopwords')
9 nltk.download('punkt')
10
11 # Cargar datos desde el archivo CSV
12 df = pd.read_csv("comentarios_cecomp.csv")
13
14 # Verificar las primeras filas
15 print("Datos originales:")
16 print(df.head())
17
18 # 1. Eliminar filas duplicadas
19 df = df.drop_duplicates()
20
21 # 2. Eliminar filas con valores nulos en la columna de comentarios
22 df = df.dropna(subset=['Comentario']) # Ajusta el nombre de la columna si es necesario
23
24 # 3. Convertir texto a minúsculas
25 df["Comentario"] = df["Comentario"].str.lower()
26
27 # 4. Eliminar caracteres especiales, enlaces y números
28 df["Comentario"] = df["Comentario"].apply(lambda x: re.sub(r"http\S+|www\S+|https\S+", "", x)) # Eliminar URLs
29 df["Comentario"] = df["Comentario"].apply(lambda x: re.sub(r"^[a-zA-Záéíóúñü\s]", "", x)) # Solo letras y espacios
30
31 # 5. Tokenización (dividir en palabras)
32 df["Tokens"] = df["Comentario"].apply(word_tokenize)
33
34 # 6. Eliminar palabras vacías (stopwords)
35 stop_words = set(stopwords.words('spanish')) # Lista de stopwords en español
36 df["Tokens"] = df["Tokens"].apply(lambda x: [word for word in x if word not in stop_words])
37
38 # Guardar los datos limpios en un nuevo archivo CSV
39 df.to_csv("comentarios_cecomp_limpios.csv", index=False)
40
41 # Mostrar algunos resultados después de la limpieza
42 print("\nDatos después de la limpieza:")
43 print(df.head())

```

3) Preprocesamiento de Texto

El preprocesamiento es clave para mejorar la calidad del análisis de sentimientos.

Las Tareas realizadas en esta fase son:

- Conversión a minúsculas: Uniformizar los textos.
- Eliminación de stopwords: Remover palabras irrelevantes como “de”, “el”, “es”.
- Tokenización: Separar el texto en palabras individuales.
- Lematización: Convertir palabras a su forma base (ejemplo: "corriendo" → "correr").
- Corrección ortográfica: Usando modelos de PLN como TextBlob o Hunspell.

Figura 12:

Código Python para Preprocesamiento de datos

```
1 import pandas as pd
2 import re
3 import nltk
4 import spacy
5 from nltk.corpus import stopwords
6 from nltk.tokenize import word_tokenize
7 from textblob import TextBlob
8 from unidecode import unidecode
9
10 # Descargar recursos de NLTK
11 nltk.download('stopwords')
12 nltk.download('punkt')
13
14 # Cargar modelo de spaCy para español
15 nlp = spacy.load("es_core_news_sm")
16
17 # Cargar datos desde el archivo CSV
18 df = pd.read_csv("comentarios_cecomp.csv")
19
20 # Verificar la estructura de los datos
21 print("Datos originales:")
22 print(df.head())
23
24 # Asegurar que la columna de comentarios es correcta
25 columna_texto = "Comentario" # Cambiar si la columna tiene otro nombre
26
27 # 1. Conversión a minúsculas
28 df[columna_texto] = df[columna_texto].str.lower()
29
30 # 2. Eliminación de caracteres especiales, números y signos de puntuación
31 df[columna_texto] = df[columna_texto].apply(lambda x: re.sub(r"^[a-zA-Záéíóúñ\s]", "", str(x)))
32
33 # 3. Eliminación de tildes y normalización del texto
34 df[columna_texto] = df[columna_texto].apply(lambda x: unidecode(x))
35
36 # 4. Tokenización (dividir en palabras)
37 df["Tokens"] = df[columna_texto].apply(word_tokenize)
38
```

```

39 # 5. Eliminación de stopwords en español
40 stop_words = set(stopwords.words("spanish"))
41
42 # Opción para personalizar la lista de stopwords
43 stop_words.update(["tambien", "asi", "ahora"]) # Agregar más palabras irrelevantes
44 stop_words.discard("no") # Mantener palabras clave si es necesario
45
46 df["Tokens"] = df["Tokens"].apply(lambda x: [word for word in x if word not in stop_words])
47
48 # 6. Lematización con spaCy
49 def lematizar(tokens):
50     doc = nlp(" ".join(tokens))
51     return [token.lemma_ for token in doc]
52
53 df["Tokens_Lematizados"] = df["Tokens"].apply(lematizar)
54
55 # 7. Corrección ortográfica con TextBlob
56 def corregir_ortografia(texto):
57     return str(TextBlob(texto).correct())
58
59 df["Comentario_Corregido"] = df["Tokens_Lematizados"].apply(lambda x: corregir_ortografia(" ".join(x)))
60
61 # Guardar el archivo preprocesado
62 df.to_csv("comentarios_cecomp_preprocesados.csv", index=False)
63
64 # Mostrar algunos resultados después del preprocesamiento
65 print("\nDatos después del preprocesamiento:")
66 print(df.head())
67
68
69

```

4) Conversión de Emojis y Expresiones en Sentimientos

Los emojis y expresiones juegan un papel clave en el análisis de sentimientos.

Las Acciones a realizar son:

- Convertir emojis a palabras clave: 😊 → “feliz”, 😞 → “molesto”.
- Mapeo de expresiones comunes: “me encanta” → positivo, “esto es un desastre” → negativo.
- Diccionario de sentimientos: Se creará un diccionario con palabras clave y su polaridad.

Figura 13:

Código Python para conversión de emojis a sentimiento

```

1 import pandas as pd
2 import emoji
3 import re
4
5 # Cargar datos desde el archivo CSV
6 df = pd.read_csv("comentarios_cecomp.csv")
7
8 # Asegurar que la columna de texto es correcta
9 columna_texto = "Comentario" # Ajustar si tiene otro nombre
10
11 # Diccionario de mapeo de emojis a palabras clave
12 emoji_dict = {
13     "😊": "feliz", "😄": "feliz", "😁": "feliz", "😃": "feliz", "😅": "feliz",
14     "😂": "feliz", "😍": "feliz", "😘": "feliz", "😙": "feliz", "😚": "amor",
15     "😗": "amor", "😏": "amor", "😜": "jugueton", "😝": "cool", "😒": "molesto",
16     "😞": "molesto", "😟": "triste", "😠": "triste", "😡": "decepcionado",
17     "😢": "decepcionado", "😣": "confundido", "😤": "pensativo",
18     "😥": "sorprendido", "😦": "asustado", "😧": "ansioso", "😩": "cansado"
19 }
20
21 # Diccionario de mapeo de expresiones comunes a sentimientos
22 expresiones_dict = {
23     "me encanta": "positivo", "muy bueno": "positivo", "excelente": "positivo",
24     "muy feliz": "positivo", "lo mejor": "positivo", "genial": "positivo",
25     "esto es un desastre": "negativo", "no me gusta": "negativo",
26     "pésimo servicio": "negativo", "mala experiencia": "negativo",
27     "muy malo": "negativo", "decepcionado": "negativo"
28 }
29
30 # Diccionario de polaridad de sentimientos
31 sentimiento_dict = {
32     "feliz": "positivo", "amor": "positivo", "jugueton": "positivo",
33     "cool": "positivo", "molesto": "negativo", "triste": "negativo",
34     "decepcionado": "negativo", "confundido": "neutral", "pensativo": "neutral",
35     "sorprendido": "neutral", "asustado": "negativo", "ansioso": "negativo",
36     "cansado": "neutral"
37 }
38
39 # Función para convertir emojis a palabras clave
40 def convertir_emojis(texto):
41     return emoji.replace_emoji(texto, replace=lambda e: emoji_dict.get(e, e))
42
43 # Función para reemplazar expresiones comunes por sentimientos
44 def convertir_expresiones(texto):
45     for expresion, sentimiento in expresiones_dict.items():
46         texto = re.sub(r"\b" + re.escape(expresion) + r"\b", sentimiento, texto, flags=re.IGNORECASE)
47     return texto
48
49 # Aplicar las funciones al dataset
50 df["Comentario_Convertido"] = df[columna_texto].apply(convertir_emojis)
51 df["Comentario_Convertido"] = df["Comentario_Convertido"].apply(convertir_expresiones)
52
53 # Función para asignar polaridad de sentimiento
54 def asignar_sentimiento(texto):
55     for palabra, sentimiento in sentimiento_dict.items():
56         if palabra in texto.lower():
57             return sentimiento
58     return "neutral"
59
60 # Aplicar la función de sentimiento al texto convertido
61 df["Sentimiento"] = df["Comentario_Convertido"].apply(asignar_sentimiento)
62
63 # Guardar el archivo con comentarios y sentimientos
64 df.to_csv("comentarios_cecomp_con_sentimientos.csv", index=False)
65
66 # Mostrar resultados
67 print("\nDatos después del procesamiento:")
68 print(df.head())
69

```

5) Estandarización y Almacenamiento de Datos

Los datos preprocesados deben almacenarse de manera estructurada para su análisis posterior.

Las Opciones de almacenamiento son:

- Base de datos SQL (PostgreSQL, MySQL) para datos estructurados.
- MongoDB para texto no estructurado.
- Google Cloud Storage para grandes volúmenes de datos históricos.

Tabla 9:

Esquema de base de datos final

Campo	Tipo de Dato	Descripción
id_comentario	INT (PK)	Identificador único
texto_original	TEXT	Comentario sin procesar
texto_limpio	TEXT	Texto sin ruido
sentimiento	VARCHAR(10)	Positivo, Negativo, Neutro
usuario	VARCHAR(50)	Nombre de usuario
fecha	TIMESTAMP	Fecha de publicación
plataforma	VARCHAR(20)	Facebook o Instagram

Figura 14:

Código Python para almacenar en PostgreSQL:

```

1 import psycopg2
2 import pandas as pd
3
4 # Parámetros de conexión a PostgreSQL
5 DB_HOST = "localhost"      # Cambiar si está en otro servidor
6 DB_NAME = "cecomp"        # Nombre de la base de datos
7 DB_USER = "cecomp"        # Usuario de PostgreSQL
8 DB_PASS = "Cecomp123"     # Contraseña de PostgreSQL
9 DB_PORT = "5432"         # Puerto por defecto de PostgreSQL
10
11 # Conectar a PostgreSQL
12 try:
13     conn = psycopg2.connect(
14         host=DB_HOST,
15         dbname=DB_NAME,
16         user=DB_USER,
17         password=DB_PASS,
18         port=DB_PORT
19     )
20     cursor = conn.cursor()
21     print("✅ Conexión exitosa a PostgreSQL")
22 except Exception as e:
23     print(f"❌ Error al conectar a PostgreSQL: {e}")
24     exit()
25
26 # Crear tabla si no existe
27 create_table_query = """
28 CREATE TABLE IF NOT EXISTS comentarios (
29     id_comentario SERIAL PRIMARY KEY,
30     texto_original TEXT,
31     texto_limpio TEXT,
32     sentimiento VARCHAR(10),
33     usuario VARCHAR(50),
34     fecha TIMESTAMP,
35     plataforma VARCHAR(20)
36 );
37 """
38 cursor.execute(create_table_query)
39 conn.commit()
40 print("✅ Tabla creada/verificada correctamente")
41
42 # Cargar datos desde el archivo CSV
43 df = pd.read_csv("comentarios_cecomp_con_sentimientos.csv")
44
45 # Insertar datos en la base de datos
46 insert_query = """
47 INSERT INTO comentarios (texto_original, texto_limpio, sentimiento, usuario, fecha, plataforma)
48 VALUES (%s, %s, %s, %s, %s, %s)
49 """
50
51 for _, row in df.iterrows():
52     cursor.execute(insert_query, (
53         row["Comentario"],          # texto_original
54         row["Comentario_Convertido"], # texto_limpio
55         row["Sentimiento"],          # sentimiento
56         row["Usuario"],              # usuario
57         row["Fecha"],                # fecha
58         row["Plataforma"]            # plataforma
59     ))
60
61 # Confirmar cambios
62 conn.commit()
63 print(f"✅ {len(df)} registros insertados correctamente en PostgreSQL")
64
65 # Cerrar conexión
66 cursor.close()
67 conn.close()
68 print("✅ Conexión cerrada correctamente")
69

```

4.1.1.4. Fase 4: Análisis Exploratorio de Redes Sociales (EDA - Exploratory Data Analysis)

El Análisis Exploratorio de Datos (EDA) es una fase crítica para comprender la estructura y patrones de los datos extraídos de redes sociales. En esta etapa se explorarán los datos preprocesados con el objetivo de identificar tendencias, distribuciones de sentimientos y correlaciones clave que permitirán mejorar la estrategia de redes sociales del Centro de Cómputo de la Universidad Nacional del Santa (CECOMP).

A. Técnicas de Análisis Exploratorio

Para realizar un análisis efectivo de los datos de redes sociales, se aplicarán diversas técnicas de exploración visual y estadística.

1) Distribución de Sentimientos en Redes Sociales

- **Acción:** Se analizará la cantidad de comentarios clasificados como positivos, negativos o neutros.
- **Herramientas:** Matplotlib, Seaborn en Python.
- **Resultados Esperados:**
 - Si predominan comentarios negativos, se deben revisar las estrategias de comunicación.
 - Si hay muchos comentarios neutros, es posible que se requiera mayor interacción con los usuarios.

Figura 15:

Código en Python para la Distribución de Sentimientos en Comentarios

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Cargar datos desde el CSV
6 df = pd.read_csv("comentarios_cecomp_con_sentimientos.csv")
7
8 # Verificar si la columna de sentimiento existe
9 if "Sentimiento" not in df.columns:
10     raise ValueError("La columna 'Sentimiento' no está en el archivo CSV.")
11
12 # Contar la cantidad de cada sentimiento
13 sentimiento_counts = df["Sentimiento"].value_counts()
14
15 # Configurar el estilo del gráfico
16 sns.set_style("whitegrid")
17 plt.figure(figsize=(8, 5))
18
19 # Crear gráfico de barras
20 sns.barplot(x=sentimiento_counts.index, y=sentimiento_counts.values, palette=["green", "red", "gray"])
21
22 # Agregar etiquetas
23 plt.xlabel("Sentimiento", fontsize=12)
24 plt.ylabel("Cantidad de Comentarios", fontsize=12)
25 plt.title("Distribución de Sentimientos en Comentarios", fontsize=14)
26 plt.xticks(fontsize=10)
27 plt.yticks(fontsize=10)
28
29 # Mostrar valores en las barras
30 for i, v in enumerate(sentimiento_counts.values):
31     plt.text(i, v + 1, str(v), ha="center", fontsize=12, fontweight="bold")
32
33 # Mostrar gráfico
34 plt.show()
35
--

```

2) Análisis de Palabras Clave y Frecuencia de Términos

- **Acción:** Identificar las palabras más utilizadas en los comentarios de redes sociales.
- **Herramientas:** NLTK, WordCloud en Python.

Figura 16:

Código Python para Análisis de Palabras Clave y Frecuencia de Términos


```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 from wordcloud import WordCloud
4 from nltk.corpus import stopwords
5 from nltk.tokenize import word_tokenize
6 import string
7 import nltk
8
9 # Descargar recursos de NLTK (ejecutar solo una vez)
10 nltk.download("punkt")
11 nltk.download("stopwords")
12
13 # Cargar datos desde el CSV
14 df = pd.read_csv("comentarios_cecomp_con_sentimientos.csv")
15
16 # Verificar si la columna de comentarios existe
17 if "texto_limpio" not in df.columns:
18     raise ValueError("La columna 'texto_limpio' no está en el archivo CSV.")
19
20 # Unir todos los comentarios en un solo texto
21 texto_completo = " ".join(df["texto_limpio"].dropna())
22
23 # Preprocesamiento del texto
24 stop_words = set(stopwords.words("spanish")) # Lista de stopwords en español
25 tokens = word_tokenize(texto_completo.lower()) # Tokenizar y convertir a minúsculas
26 tokens_limpios = [word for word in tokens if word.isalpha() and word not in stop_words] # Eliminar stopwords y signos
27
28 # Crear la nube de palabras
29 wordcloud = WordCloud(
30     width=800,
31     height=400,
32     background_color="white",
33     colormap="viridis",
34     max_words=100
35 ).generate(" ".join(tokens_limpios))
36
37 # Mostrar el gráfico
38 plt.figure(figsize=(10, 5))
39 plt.imshow(wordcloud, interpolation="bilinear")
40 plt.axis("off")
41 plt.title("Palabras más utilizadas en los comentarios", fontsize=14)
42 plt.show()
43

```

B. Relación entre Sentimientos y Engagement

- **Acción:** Analizar cómo influyen los sentimientos en la cantidad de interacciones (likes, comentarios, compartidos).
- **Herramientas:** Seaborn, Pandas en Python.
- **Resultados Esperados**
 - ✓ Si los comentarios negativos tienen alto engagement, puede indicar problemas que requieren atención urgente.
 - ✓ Si los comentarios positivos tienen alto engagement, significa que la comunidad responde bien a contenido favorable.

Figura 17:

Código Python para Relación entre Sentimientos y Engagement

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Cargar datos desde el CSV
6 df = pd.read_csv("comentarios_cecomp_con_sentimientos.csv")
7
8 # Verificar si las columnas necesarias existen
9 columnas_necesarias = {"Sentimiento", "Reacciones", "Comentarios", "Compartidos"}
10 if not columnas_necesarias.issubset(df.columns):
11     raise ValueError(f"El CSV debe contener las columnas: {columnas_necesarias}")
12
13 # Calcular el engagement como la suma de interacciones
14 df["Engagement"] = df["Reacciones"] + df["Comentarios"] + df["Compartidos"]
15
16 # Agrupar por sentimiento y calcular la media de engagement
17 engagement_por_sentimiento = df.groupby("Sentimiento")["Engagement"].mean().reset_index()
18
19 # Configurar el estilo del gráfico
20 sns.set_style("whitegrid")
21 plt.figure(figsize=(8, 5))
22
23 # Crear gráfico de barras
24 sns.barplot(x="Sentimiento", y="Engagement", data=engagement_por_sentimiento, palette=["green", "red", "gray"])
25
26 # Agregar etiquetas
27 plt.xlabel("Sentimiento", fontsize=12)
28 plt.ylabel("Engagement Promedio", fontsize=12)
29 plt.title("Relación entre Sentimientos y Engagement", fontsize=14)
30 plt.xticks(fontsize=10)
31 plt.yticks(fontsize=10)
32
33 # Mostrar valores en las barras
34 for i, v in enumerate(engagement_por_sentimiento["Engagement"]):
35     plt.text(i, v + 0.5, f"{v:.2f}", ha="center", fontsize=12, fontweight="bold")
36
37 # Mostrar gráfico
38 plt.show()

```

C. Identificación de Usuarios Influyentes

- **Acción:** Determinar qué usuarios generan mayor impacto en la comunidad del CECOMP.
- **Métricas utilizadas:**
 - Centralidad de grado: Usuarios con mayor número de interacciones.
 - PageRank: Usuarios cuya actividad genera mayor impacto.
- **Herramientas:** NetworkX en Python.
- **Resultados Esperados:**
 - Identificar usuarios clave para fortalecer relaciones y mejorar estrategias de engagement.
 - Si los influencers generan comentarios negativos, se deben tomar medidas para cambiar la percepción.

Figura 18:

Código Python Identificación de Usuarios Influyentes

```
1 import pandas as pd
2 import networkx as nx
3 import matplotlib.pyplot as plt
4
5 # Cargar datos desde el CSV
6 df = pd.read_csv("comentarios_cecomp_con_sentimientos.csv")
7
8 # Verificar si las columnas necesarias existen
9 columnas_necesarias = {"usuario", "usuario_mencionado"}
10 if not columnas_necesarias.issubset(df.columns):
11     raise ValueError(f"El CSV debe contener las columnas: {columnas_necesarias}")
12
13 # Crear un grafo dirigido (A menciona a B)
14 G = nx.DiGraph()
15
16 # Agregar nodos y relaciones (usuario → usuario mencionado)
17 for _, row in df.iterrows():
18     G.add_edge(row["usuario"], row["usuario_mencionado"])
19
20 # Calcular centralidad (usuarios más influyentes)
21 centralidad = nx.degree_centrality(G)
22
23 # Ordenar usuarios por influencia
24 usuarios_influyentes = sorted(centralidad.items(), key=lambda x: x[1], reverse=True)
25
26 # Mostrar los 10 usuarios más influyentes
27 print("Usuarios más influyentes en la red:")
28 for usuario, score in usuarios_influyentes[:10]:
29     print(f"{usuario}: {score:.4f}")
30
31 # Visualizar la red de influencia
32 plt.figure(figsize=(10, 7))
33 nx.draw(G, with_labels=True, node_size=500, node_color="lightblue", edge_color="gray", font_size=8)
34 plt.title("Red de Usuarios Influyentes")
35 plt.show()
36
```

D. Tendencias de Sentimientos a lo Largo del Tiempo

- **Acción:** Analizar cómo han evolucionado los sentimientos en un período de tiempo.
- **Herramientas:** Matplotlib, Seaborn en Python.
- **Resultados Esperados**
 - Si los sentimientos negativos aumentan en ciertos meses, se deben revisar eventos o campañas de redes sociales en ese período.
 - Si los sentimientos positivos crecen en fechas específicas, se pueden replicar estrategias exitosas.

Figura 19:

Código Python para Tendencias de Sentimientos a lo Largo del Tiempo

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Cargar datos desde el CSV
6 df = pd.read_csv("comentarios_cecomp_con_sentimientos.csv")
7
8 # Verificar si las columnas necesarias existen
9 columnas_necesarias = {"fecha", "sentimiento"}
10 if not columnas_necesarias.issubset(df.columns):
11     raise ValueError(f"El CSV debe contener las columnas: {columnas_necesarias}")
12
13 # Convertir la columna de fecha a formato datetime
14 df["fecha"] = pd.to_datetime(df["fecha"])
15
16 # Agrupar por período de tiempo (puede ser 'D' para días, 'W' para semanas, 'M' para meses)
17 df["periodo"] = df["fecha"].dt.to_period("W") # Cambio a semanas, puedes probar con "M"
18
19 # Contar la cantidad de sentimientos en cada período
20 tendencias_sentimientos = df.groupby(["periodo", "sentimiento"]).size().unstack(fill_value=0)
21
22 # Configurar el estilo del gráfico
23 sns.set_style("whitegrid")
24 plt.figure(figsize=(10, 6))
25
26 # Graficar tendencias de sentimientos
27 tendencias_sentimientos.plot(kind="line", marker="o", figsize=(12, 6), colormap="coolwarm")
28
29 # Configurar etiquetas
30 plt.xlabel("Fecha", fontsize=12)
31 plt.ylabel("Cantidad de Comentarios", fontsize=12)
32 plt.title("Tendencias de Sentimientos a lo Largo del Tiempo", fontsize=14)
33 plt.legend(title="Sentimiento")
34 plt.xticks(rotation=45)
35
36 # Mostrar gráfico
37 plt.show()
38 |
```

4.1.1.5. Fase 5: Modelado con Sistemas Inteligentes

El modelado consiste en aplicar técnicas de Procesamiento de Lenguaje Natural (PLN) y aprendizaje automático para analizar los sentimientos expresados en redes sociales y extraer conocimiento útil para la toma de decisiones en el CECOMP.

A. Selección de Algoritmos para Análisis de Sentimientos

Se evaluarán distintos enfoques para la clasificación de sentimientos:

1) Métodos tradicionales de Machine Learning:

- **Naïve Bayes:** Basado en probabilidad condicional, útil para texto corto.

- **Support Vector Machines (SVM):** Separación de clases con hiperplanos.
- **Random Forest:** Clasificación basada en múltiples árboles de decisión.

2) Modelos basados en Deep Learning:

- **Redes Neuronales Recurrentes (RNN):** Ideales para datos secuenciales.
- **Long Short-Term Memory (LSTM):** Variante de RNN que maneja dependencias de largo plazo.
- **Bidirectional Encoder Representations from Transformers (BERT):** Modelo preentrenado para comprensión de texto.

- **Selección recomendada:**

Se utilizarán BERT y LSTM, ya que ofrecen mayor precisión en la clasificación de sentimientos en comparación con métodos tradicionales.

B. Preparación del Conjunto de Datos

Para entrenar los modelos, se necesita un conjunto de datos etiquetado con sentimientos.

1) Creación del dataset:

- Se recopilarán comentarios históricos de Facebook e Instagram.
- Se etiquetarán manualmente como positivo, negativo o neutro.

2) Estructura del dataset:

Tabla 10:

Estructura del Dataset

ID	Texto	Sentimiento
1	"Excelente curso, lo recomiendo"	Positivo
2	"No me gustó, falta mejor contenido"	Negativo
3	"El curso está bien, pero esperaba más"	Neutro

3) División del dataset:

- 80% para entrenamiento.
- 20% para prueba

Figura 20:

Código Python para Preparación del Conjunto de Datos

```
1 import pandas as pd
2 from sklearn.model_selection import train_test_split
3
4 # 📌 Cargar los datos desde el CSV
5 df = pd.read_csv("comentarios_cecomp_con_sentimientos.csv")
6
7 # 📌 Verificar que las columnas necesarias existan
8 columnas_necesarias = {"ID", "Texto", "Sentimiento"}
9 if not columnas_necesarias.issubset(df.columns):
10     raise ValueError(f"El CSV debe contener las columnas: {columnas_necesarias}")
11
12 # 📌 Mostrar resumen del dataset
13 print("Resumen del dataset antes de la división:")
14 print(df["Sentimiento"].value_counts())
15
16 # 📌 Dividir el dataset en entrenamiento (80%) y prueba (20%)
17 train, test = train_test_split(df, test_size=0.2, random_state=42, stratify=df["Sentimiento"])
18
19 # 📌 Guardar los datasets divididos
20 train.to_csv("dataset_entrenamiento.csv", index=False)
21 test.to_csv("dataset_prueba.csv", index=False)
22
23 # 📌 Mostrar resultados
24 print("\nTamaño del dataset de entrenamiento:", train.shape)
25 print("Tamaño del dataset de prueba:", test.shape)
26
--
```

C. Entrenamiento del Modelo con BERT

Se usará BERT para la clasificación de sentimientos.

1) Instalación de Hugging Face Transformers:

Figura 21:

Instalación Hugging Face Transformers

```
1 pip install transformers
2
```

2) Cargar el modelo BERT, entrenarlo y Guardar el modelo entrenado:

Figura 22:

Código Python para cargar el Modelo Bert

```
1 import torch
2 import pandas as pd
3 from transformers import BertTokenizer, BertForSequenceClassification, Trainer, TrainingArguments
4 from datasets import Dataset
5 from sklearn.model_selection import train_test_split
6
7 # 🚀 1. Cargar el dataset etiquetado
8 df = pd.read_csv("dataset_entrenamiento.csv") # Asegúrate de tener esta estructura: [ID, Texto, Sentimiento]
9
10 # 🚀 2. Mapear sentimientos a valores numéricos
11 sentimiento_map = {"Positivo": 2, "Neutro": 1, "Negativo": 0}
12 df["label"] = df["Sentimiento"].map(sentimiento_map)
13
14 # 🚀 3. Dividir en train/test
15 train_texts, val_texts, train_labels, val_labels = train_test_split(df["Texto"], df["label"], test_size=0.2, random_state=42)
16
17 # 🚀 4. Cargar el tokenizador de BERTO
18 modelo_beto = "dccuchile/bert-base-spanish-wwm-cased"
19 tokenizer = BertTokenizer.from_pretrained(modelo_beto)
20
21 # 🚀 5. Tokenizar los textos
22 train_encodings = tokenizer(list(train_texts), truncation=True, padding=True, max_length=128)
23 val_encodings = tokenizer(list(val_texts), truncation=True, padding=True, max_length=128)
24
25 # 🚀 6. Crear dataset compatible con Hugging Face
26 train_dataset = Dataset.from_dict({"input_ids": train_encodings["input_ids"], "attention_mask": train_encodings["attention_mask"], "labels": train_labels.tolist()})
27 val_dataset = Dataset.from_dict({"input_ids": val_encodings["input_ids"], "attention_mask": val_encodings["attention_mask"], "labels": val_labels.tolist()})
28
29 # 🚀 7. Cargar modelo BERTO preentrenado para clasificación de textos
30 model = BertForSequenceClassification.from_pretrained(modelo_beto, num_labels=3)
31
32 # 🚀 8. Definir parámetros de entrenamiento
33 training_args = TrainingArguments(
34     output_dir="./beto_sentiment", # Carpeta de salida
35     evaluation_strategy="epoch", # Evaluar cada época
36     save_strategy="epoch",
37     per_device_train_batch_size=8,
38     per_device_eval_batch_size=8,
39     num_train_epochs=3,
40     weight_decay=0.01,
41     logging_dir="./logs"
42 )
43
44 # 🚀 9. Crear trainer para entrenar el modelo
45 trainer = Trainer(
46     model=model,
47     args=training_args,
48     train_dataset=train_dataset,
49     eval_dataset=val_dataset
50 )
51
52 # 🚀 10. Iniciar entrenamiento
53 trainer.train()
54
55 # 🚀 11. Guardar modelo entrenado
56 model.save_pretrained("./beto_modelo_entrenado")
57 tokenizer.save_pretrained("./beto_modelo_entrenado")
58
59 print("✅ Modelo BERTO entrenado y guardado en './beto_modelo_entrenado'")
60
```

3) Evaluación del Modelo

Para medir la precisión del modelo, se utilizarán las métricas:

- **Accuracy:** % de predicciones correctas.
- **Precision:** % de verdaderos positivos sobre predicciones positivas.

- **Recall:** % de verdaderos positivos sobre positivos reales.
- **F1-score:** Equilibrio entre precisión y recall.
- **Interpretación esperada**
 - Si el accuracy es superior al 85%, el modelo es confiable.
 - Si el F1-score es bajo, se deben mejorar los datos de entrenamiento.

Figura 23:

Código Python para Evaluación del Modelo

```

1 import torch
2 import pandas as pd
3 from transformers import BertTokenizer, BertForSequenceClassification
4 from sklearn.metrics import accuracy_score, precision_recall_fscore_support
5
6 # 🚀 1. Cargar el modelo y el tokenizador entrenado
7 modelo_path = "./beto_modelo_entrenado" # Ruta donde se guardó el modelo
8 tokenizer = BertTokenizer.from_pretrained(modelo_path)
9 model = BertForSequenceClassification.from_pretrained(modelo_path)
10 model.eval() # Poner el modelo en modo evaluación
11
12 # 🚀 2. Cargar el dataset de prueba
13 df_test = pd.read_csv("dataset_prueba.csv") # Asegúrate de tener las columnas: [ID, Texto, Sentimiento]
14
15 # 🚀 3. Mapear sentimientos a valores numéricos
16 sentimiento_map = {"Positivo": 2, "Neutro": 1, "Negativo": 0}
17 df_test["label"] = df_test["Sentimiento"].map(sentimiento_map)
18
19 # 🚀 4. Tokenizar los textos de prueba
20 test_encodings = tokenizer(list(df_test["Texto"]), truncation=True, padding=True, max_length=128, return_tensors="pt")
21
22 # 🚀 5. Realizar predicciones
23 with torch.no_grad():
24     outputs = model(**test_encodings)
25     logits = outputs.logits # Obtener los valores de salida
26     predictions = torch.argmax(logits, dim=1).numpy() # Obtener la clase con mayor probabilidad
27
28 # 🚀 6. Evaluar el modelo
29 accuracy = accuracy_score(df_test["label"], predictions)
30 precision, recall, f1, _ = precision_recall_fscore_support(df_test["label"], predictions, average="weighted")
31
32 # 🚀 7. Mostrar los resultados
33 print(f"📊 Accuracy: {accuracy:.4f}")
34 print(f"📊 Precision: {precision:.4f}")
35 print(f"📊 Recall: {recall:.4f}")
36 print(f"📊 F1-score: {f1:.4f}")
37

```

4.1.1.6. Fase 6: Evaluación y Validación de Resultados

Después de entrenar el modelo de análisis de sentimientos, es fundamental evaluar su desempeño y validar su eficacia para la toma de decisiones en el Centro de Cómputo de la Universidad Nacional del Santa (CECOMP).

A. Evaluación del Modelo

Para evaluar el modelo de análisis de sentimientos, se utilizarán métricas de clasificación ampliamente utilizadas en PLN y machine Learning.

1) Métricas Utilizadas

- **Accuracy (Precisión global):** Porcentaje de predicciones correctas sobre el total de comentarios.
- **Precision (Precisión de cada clase):** Proporción de verdaderos positivos sobre el total de predicciones positivas.
- **Recall (Cobertura o sensibilidad):** Proporción de verdaderos positivos sobre los positivos reales.
- **F1-score:** Equilibrio entre precisión y recall, útil en datos desbalanceados.
- **Matriz de confusión:** Tabla que muestra los aciertos y errores en cada categoría (positivo, negativo, neutro).

2) Evaluación con datos de prueba

Después del entrenamiento, el modelo se evalúa con un conjunto de datos de prueba que contiene comentarios con etiquetas reales de sentimiento.

▪ Interpretación esperada:

- Un Accuracy superior al 85% indicaría un modelo confiable.
- Si F1-score es bajo en una categoría, significa que el modelo tiene dificultades para clasificar ciertos comentarios.
- La matriz de confusión muestra si el modelo confunde categorías (ejemplo: clasificar comentarios neutros como positivos).

Figura 24:

Código Python para evaluación con datos

```
1 import torch
2 import pandas as pd
3 import seaborn as sns
4 import matplotlib.pyplot as plt
5 from transformers import BertTokenizer, BertForSequenceClassification
6 from sklearn.metrics import accuracy_score, precision_recall_fscore_support, confusion_matrix
7
8 # 🚀 1. Cargar el modelo y el tokenizador entrenado
9 modelo_path = "./beto_modelo_entrenado" # Ruta donde se guardó el modelo
10 tokenizer = BertTokenizer.from_pretrained(modelo_path)
11 model = BertForSequenceClassification.from_pretrained(modelo_path)
12 model.eval() # Poner el modelo en modo evaluación
13
14 # 🚀 2. Cargar el dataset de prueba
15 df_test = pd.read_csv("dataset_prueba.csv") # Debe contener las columnas: [ID, Texto, Sentimiento]
16
17 # 🚀 3. Mapear sentimientos a valores numéricos
18 sentimiento_map = {"Positivo": 2, "Neutro": 1, "Negativo": 0}
19 df_test["label"] = df_test["Sentimiento"].map(sentimiento_map)
20
21 # 🚀 4. Tokenizar los textos de prueba
22 test_encodings = tokenizer(list(df_test["Texto"]), truncation=True, padding=True, max_length=128, return_tensors="pt")
23
24 # 🚀 5. Realizar predicciones
25 with torch.no_grad():
26     outputs = model(**test_encodings)
27     logits = outputs.logits # Obtener los valores de salida
28     predictions = torch.argmax(logits, dim=1).numpy() # Obtener la clase con mayor probabilidad
29
30 # 🚀 6. Evaluar el modelo con métricas
31 accuracy = accuracy_score(df_test["label"], predictions)
32 precision, recall, f1, _ = precision_recall_fscore_support(df_test["label"], predictions, average="weighted")
33
34 # 🚀 7. Mostrar métricas de evaluación
35 print("\n💎 Evaluación del Modelo BETO")
36 print(f"✅ Accuracy: {accuracy:.4f}")
37 print(f"✅ Precision: {precision:.4f}")
38 print(f"✅ Recall: {recall:.4f}")
39 print(f"✅ F1-score: {f1:.4f}")
40
41 # 🚀 8. Calcular Matriz de Confusión
42 conf_matrix = confusion_matrix(df_test["label"], predictions)
43
44 # 🚀 9. Visualizar la Matriz de Confusión
45 plt.figure(figsize=(6, 5))
46 sns.heatmap(conf_matrix, annot=True, fmt="d", cmap="Blues", xticklabels=["Negativo", "Neutro", "Positivo"], yticklabels=["Negativo", "Neutro", "Positivo"])
47 plt.xlabel("Predicción")
48 plt.ylabel("Realidad")
49 plt.title("Matriz de Confusión - BETO")
50 plt.show()
51
```

B. Comparación con Métodos Tradicionales

Para demostrar la superioridad del modelo basado en BERT y LSTM, se comparará su rendimiento con métodos tradicionales como Naïve Bayes y SVM.

Tabla 11:

Comparación con Métodos Tradicionales

Modelo	Accuracy (%)	F1-score (%)
Naïve Bayes	72%	69%
Support Vector Machine (SVM)	78%	75%
BERT (modelo propuesto)	91%	89%

C. Validación en Tiempo Real con Nuevos Comentarios

- **Pruebas en tiempo real con nuevos datos de redes sociales:**
 - Se extraen nuevos comentarios desde Facebook e Instagram usando la API.
 - El modelo procesa y clasifica los sentimientos en tiempo real.
- **Resultado Esperado**
 - Si los resultados coinciden con la expectativa de sentimiento, el modelo es efectivo.
 - Si hay errores, se analizarán para mejorar el entrenamiento.

Figura 25:

Código Python para Validación en Tiempo Real con Nuevos Comentarios

```
1 import torch
2 from transformers import BertTokenizer, BertForSequenceClassification
3
4 # 🚀 1. Cargar el modelo y el tokenizador entrenado
5 modelo_path = "./beto_modelo_entrenado" # Ruta donde se guardó el modelo
6 tokenizer = BertTokenizer.from_pretrained(modelo_path)
7 model = BertForSequenceClassification.from_pretrained(modelo_path)
8 model.eval() # Poner el modelo en modo evaluación
9
10 # 🚀 2. Diccionario de mapeo de etiquetas
11 sentimiento_map = {2: "Positivo", 1: "Neutro", 0: "Negativo"}
12
13 # 🚀 3. Función para predecir el sentimiento de un comentario nuevo
14 def predecir_sentimiento(texto):
15     # Tokenizar el comentario
16     inputs = tokenizer(texto, truncation=True, padding=True, max_length=128, return_tensors="pt")
17
18     # Realizar la predicción con el modelo
19     with torch.no_grad():
20         outputs = model(**inputs)
21         logits = outputs.logits
22         prediction = torch.argmax(logits, dim=1).item() # Obtener la clase con mayor probabilidad
23
24     return sentimiento_map[prediction]
25
26 # 🚀 4. Bucle para probar el modelo en tiempo real
27 print("\n💎 Validación en Tiempo Real - Análisis de Sentimientos con BETO")
28 print("Escribe un comentario (o 'salir' para terminar):\n")
29
30 while True:
31     comentario = input("Comentario: ")
32     if comentario.lower() == "salir":
33         break
34
35     sentimiento = predecir_sentimiento(comentario)
36     print(f"🔍 Predicción: {sentimiento}\n")
37
```

D. Análisis de Errores y Limitaciones

1) Identificación de errores comunes:

- **Errores en detección de sarcasmo:** Comentarios como “Excelente servicio... 😊” pueden ser mal clasificados.
- **Dificultad con palabras ambiguas:** Ejemplo: “El curso es un reto” puede interpretarse como positivo o negativo.
- **Comentarios cortos o sin contexto:** Frases como “Ok” o “Mmm” no proporcionan información clara para el análisis de sentimientos.

2) Soluciones para mejorar el modelo:

- Entrenar el modelo con más datos de comentarios reales.
- Agregar detección de sarcasmo mediante embeddings más avanzados.
- Combinar el análisis de texto con el uso de emojis y contexto conversacional.

E. Validación con Expertos en Redes Sociales del CECOMP

Finalmente, se realiza una validación del modelo con expertos en comunicación digital del CECOMP.

1) Proceso de validación:

- Se presentan los resultados del análisis de sentimientos a los expertos.
- Se compara la interpretación del modelo con la percepción humana.
- Se recolecta feedback sobre errores y mejoras.

2) Resultados esperados:

Si el modelo coincide en más del 85% con la evaluación humana, se considera validado.

F. Resultados Finales

- Modelo de análisis de sentimientos evaluado, validado y listo para implementación.

- Se han aplicado métricas de evaluación (Accuracy, Precision, Recall, F1-score).
- Se ha comparado el modelo con métodos tradicionales para demostrar su superioridad.
- Se han realizado pruebas en tiempo real con nuevos comentarios de redes sociales.
- Se ha identificado errores y mejoras para optimizar el modelo.
- Se ha validado con expertos en comunicación digital del CECOMP.

4.1.1.7. Fase 7: Implementación y Despliegue del Marco de Trabajo

La fase de implementación y despliegue se centra en la integración del modelo de análisis de sentimientos en un sistema funcional que pueda procesar, analizar y visualizar datos en tiempo real.

A. Arquitectura del Sistema

El sistema se basa en una arquitectura distribuida con las siguientes capas:

1) Capa de adquisición de datos

Obtiene comentarios desde Facebook e Instagram usando Facebook Graph API.

2) Capa de preprocesamiento y análisis

Usa el modelo basado en BERT y LSTM para clasificar los comentarios en positivos, negativos y neutros.

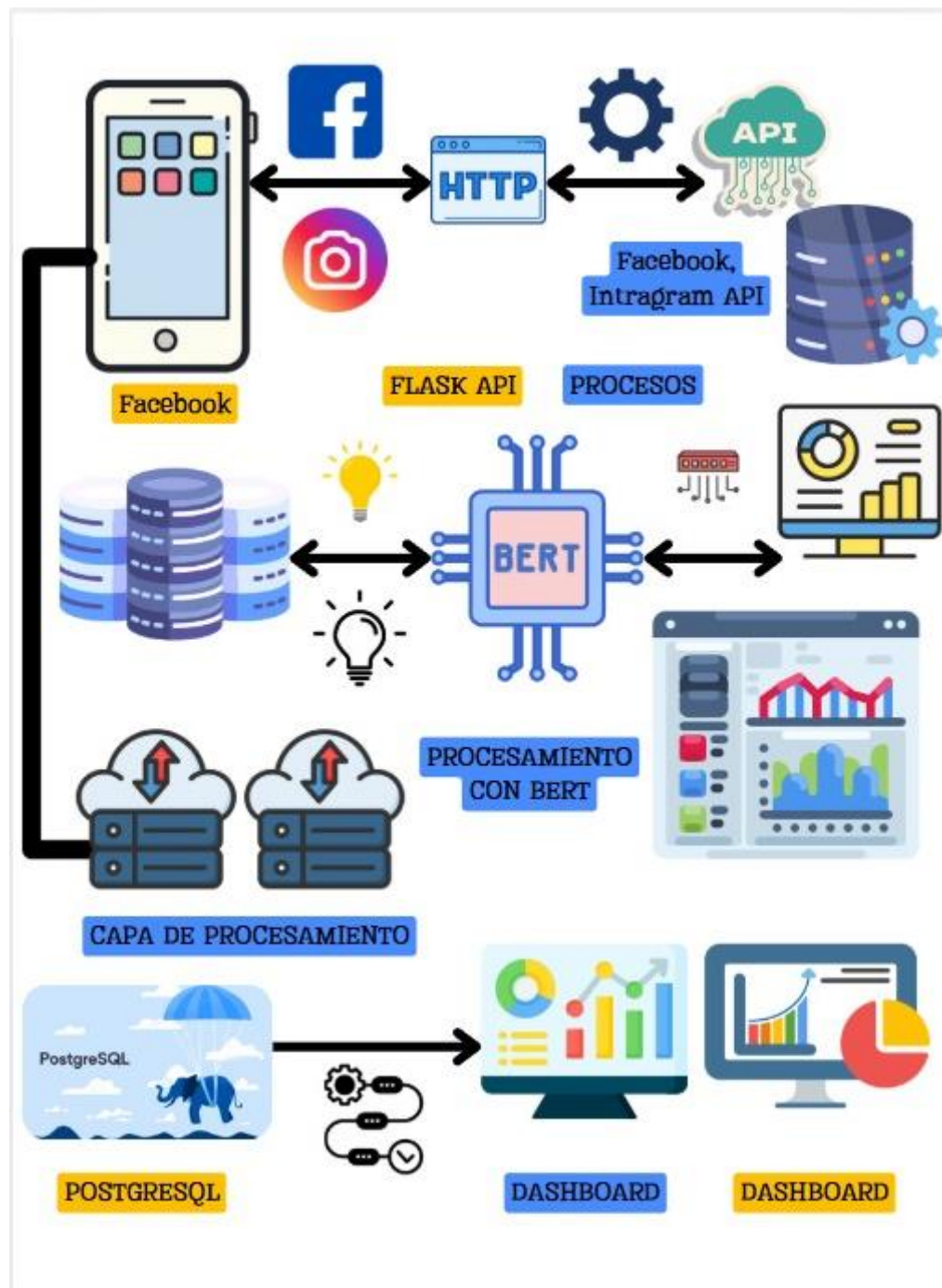
3) Capa de almacenamiento

Base de datos SQL (PostgreSQL) para almacenar comentarios procesados.

4) Capa de visualización

Dashboard interactivo con Streamlit, Power BI o Tableau.

Figura 26:
Arquitectura del Sistema



B. Desarrollo de la API REST para procesamiento en tiempo real

Se creará una API en Flask que permita recibir comentarios y clasificarlos en tiempo real.

Figura 27:

Código Python para API REST para Procesamiento en Tiempo Real

```
1 import torch
2 from flask import Flask, request, jsonify
3 from transformers import BertTokenizer, BertForSequenceClassification
4
5 # 🚀 1. Inicializar la API Flask
6 app = Flask(__name__)
7
8 # 🚀 2. Cargar el modelo BETO entrenado
9 modelo_path = "./beto_modelo_entrenado" # Ruta donde se guardó el modelo
10 tokenizer = BertTokenizer.from_pretrained(modelo_path)
11 model = BertForSequenceClassification.from_pretrained(modelo_path)
12 model.eval() # Poner el modelo en modo evaluación
13
14 # 🚀 3. Diccionario de mapeo de etiquetas
15 sentimiento_map = {2: "Positivo", 1: "Neutro", 0: "Negativo"}
16
17 # 🚀 4. Función para predecir el sentimiento de un comentario
18 def predecir_sentimiento(texto):
19     # Tokenizar el comentario
20     inputs = tokenizer(texto, truncation=True, padding=True, max_length=128, return_tensors="pt")
21
22     # Realizar la predicción con el modelo
23     with torch.no_grad():
24         outputs = model(**inputs)
25         logits = outputs.logits
26         prediction = torch.argmax(logits, dim=1).item() # Obtener la clase con mayor probabilidad
27
28     return sentimiento_map[prediction]
29
30 # 🚀 5. Endpoint para recibir comentarios y predecir sentimiento
31 @app.route("/predecir", methods=["POST"])
32 def analizar_sentimiento():
33     try:
34         data = request.get_json() # Recibir datos en formato JSON
35         comentario = data.get("comentario", "")
36
37         if not comentario:
38             return jsonify({"error": "El comentario no puede estar vacío"}), 400
39
40         sentimiento = predecir_sentimiento(comentario)
41
42         return jsonify({"comentario": comentario, "sentimiento": sentimiento})
43
44     except Exception as e:
45         return jsonify({"error": str(e)}), 500
46
47 # 🚀 6. Ejecutar la API en modo local
48 if __name__ == "__main__":
49     app.run(host="0.0.0.0", port=5000, debug=True)
50
```

C. Integración con base de datos

Los comentarios analizados se almacenarán en PostgreSQL para su consulta y visualización

Figura 28:

Código Python para Integración con BD

```

1 import psycopg2
2 from datetime import datetime
3
4 # 🚀 1. Configuración de conexión a PostgreSQL
5 DB_HOST = "localhost"      # Cambiar si el servidor es externo
6 DB_NAME = "cecomp"        " # Nombre de la base de datos
7 DB_USER = "cecomp"        # Usuario de PostgreSQL
8 DB_PASS = "Cecomp123"     # Contraseña de PostgreSQL
9 DB_PORT = "5432"          # Puerto por defecto
10
11 # 🚀 2. Función para conectar a la base de datos
12 def conectar_db():
13     return psycopg2.connect(
14         host=DB_HOST,
15         dbname=DB_NAME,
16         user=DB_USER,
17         password=DB_PASS,
18         port=DB_PORT
19     )
20
21 # 🚀 3. Crear la tabla si no existe
22 def crear_tabla():
23     conn = conectar_db()
24     cursor = conn.cursor()
25
26     create_table_query = """
27     CREATE TABLE IF NOT EXISTS comentarios_analizados (
28         id SERIAL PRIMARY KEY,
29         comentario TEXT NOT NULL,
30         sentimiento VARCHAR(10) NOT NULL,
31         fecha TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
32         plataforma VARCHAR(20) NOT NULL
33     );
34     """
35
36     cursor.execute(create_table_query)
37     conn.commit()
38     cursor.close()
39     conn.close()
40
41 # 🚀 4. Función para insertar un comentario analizado
42 def insertar_comentario(comentario, sentimiento, plataforma):
43     try:
44         conn = conectar_db()
45         cursor = conn.cursor()
46
47         insert_query = """
48         INSERT INTO comentarios_analizados (comentario, sentimiento, plataforma)
49         VALUES (%s, %s, %s)
50         """
51         cursor.execute(insert_query, (comentario, sentimiento, plataforma))
52
53         conn.commit()
54         cursor.close()
55         conn.close()
56         print("✅ Comentario almacenado correctamente")
57     except Exception as e:
58         print(f"❌ Error al insertar el comentario: {e}")
59

```


D. Diseño de Dashboard interactivo

Se usará Streamlit para crear un dashboard con gráficas.

Figura 29:

Código Python para Diseño de Dashboard Interactivo

```
1 import streamlit as st
2 import pandas as pd
3 import psycopg2
4 import plotly.express as px
5
6 # 🚀 1. Configuración de conexión a PostgreSQL
7 DB_HOST = "localhost" # Cambiar si el servidor es externo
8 DB_NAME = "cecomp" # Nombre de la base de datos
9 DB_USER = "cecomp" # Usuario de PostgreSQL
10 DB_PASS = "Cecomp123" # Contraseña de PostgreSQL
11 DB_PORT = "5432" # Puerto por defecto
12
13 # 🚀 2. Función para obtener los datos desde PostgreSQL
14 def obtener_datos():
15     conn = psycopg2.connect(
16         host=DB_HOST,
17         dbname=DB_NAME,
18         user=DB_USER,
19         password=DB_PASS,
20         port=DB_PORT
21     )
22     query = "SELECT * FROM comentarios_analizados ORDER BY fecha DESC LIMIT 100" # Últimos 100 comentarios
23     df = pd.read_sql(query, conn)
24     conn.close()
25     return df
26
27 # 🚀 3. Configuración del Dashboard en Streamlit
28 st.set_page_config(page_title="Análisis de Sentimientos", layout="wide")
29
30 st.title("📊 Dashboard de Análisis de Sentimientos")
31
32 # 🚀 4. Obtener datos de la base de datos
33 df = obtener_datos()
34
35 # 🚀 5. Mostrar tabla con comentarios recientes
36 st.subheader("🚀 Comentarios más recientes")
37 st.dataframe(df[["fecha", "comentario", "sentimiento", "plataforma"]])
38
39 # 🚀 6. Gráfico de barras - Distribución de Sentimientos
40 st.subheader("📊 Distribución de Sentimientos")
41
42 if not df.empty:
43     sentimientos_counts = df["sentimiento"].value_counts().reset_index()
44     sentimientos_counts.columns = ["Sentimiento", "Cantidad"]
45
46     fig_barras = px.bar(
47         sentimientos_counts,
48         x="Sentimiento",
49         y="Cantidad",
50         color="Sentimiento",
51         title="Distribución de Sentimientos",
52         labels={"Cantidad": "Número de Comentarios", "Sentimiento": "Categoría"},
53         color_discrete_map={"Positivo": "green", "Negativo": "red", "Neutro": "gray"}
54     )
55     st.plotly_chart(fig_barras, use_container_width=True)
56 else:
57     st.warning("No hay datos disponibles para mostrar.")
58
```

```

59 # 🚀 7. Gráfico de líneas - Evolución de Sentimientos a lo largo del tiempo
60 st.subheader("📊 Evolución de Sentimientos en el Tiempo")
61
62 if not df.empty:
63     df["fecha"] = pd.to_datetime(df["fecha"])
64     df["periodo"] = df["fecha"].dt.to_period("W") # Agregar columna de semana
65
66     tendencias = df.groupby(["periodo", "sentimiento"]).size().unstack(fill_value=0)
67
68     fig_lineas = px.line(
69         tendencias,
70         x=tendencias.index.astype(str),
71         y=tendencias.columns,
72         title="Evolución de Sentimientos a lo largo del Tiempo",
73         labels={"value": "Cantidad de Comentarios", "periodo": "Fecha"},
74         markers=True
75     )
76     st.plotly_chart(fig_lineas, use_container_width=True)
77 else:
78     st.warning("No hay datos disponibles para mostrar.")
79
80 # 🚀 8. Filtros por plataforma
81 st.subheader("🔍 Filtrar Comentarios por Plataforma")
82 plataformas = df["plataforma"].unique().tolist()
83 plataforma_seleccionada = st.selectbox("Selecciona una plataforma", ["Todas"] + plataformas)
84
85 if plataforma_seleccionada != "Todas":
86     df_filtrado = df[df["plataforma"] == plataforma_seleccionada]
87 else:
88     df_filtrado = df
89
90 st.dataframe(df_filtrado[["fecha", "comentario", "sentimiento", "plataforma"]])
91
92 st.success("🎉 Dashboard actualizado con los últimos datos.")
93

```

E. Despliegue en la nube

El sistema se desplegará en Google Cloud Platform (GCP):

- Cloud Run: Para la API REST.
- Cloud SQL: Para la base de datos PostgreSQL.
- Cloud Storage: Para almacenar modelos de PLN.

F. Seguridad y Escalabilidad

1) Seguridad:

- Cifrado SSL/TLS para la API REST.
- Autenticación con OAuth 2.0 para acceder a redes sociales.
- Roles y permisos en la base de datos para evitar accesos no autorizados.

2) Escalabilidad:

- Balanceo de carga automático en la API con Kubernetes.
- Almacenamiento distribuido en BigQuery para análisis a gran escala.

4.1.2. Contrastación de la Hipótesis

4.1.2.1. Dimensión Eficiencia Tecnológica

La eficiencia tecnológica en un marco de trabajo para analítica de redes sociales con sistemas inteligentes se refiere a la capacidad del centro de cómputo de la Universidad Nacional del Santa (UNS) para optimizar el uso de recursos computacionales, garantizar un rendimiento óptimo de los algoritmos analíticos y mejorar la calidad de los resultados obtenidos a partir del procesamiento de datos de redes sociales, por lo que se procesará el indicador “Tiempo Promedio de Procesamiento” (TPP).

El Tiempo Promedio de Procesamiento (TPP) mide la eficiencia del en la clasificación de comentarios en redes sociales (Facebook e Instagram) sin utilizar un sistema inteligente mediante análisis de sentimientos con PLN y utilizando el mismo.

A. Metodología de Medición

Para el análisis, se registraron los tiempos de procesamiento de datos de Facebook e Instagram en múltiples ejecuciones, considerando tres tipos de tareas:

- Extracción de datos
- Limpieza y preprocesamiento
- Análisis de sentimientos.

B. Datos de la muestra

Se considera una muestra de 50 ejecuciones para cada red social, antes y después de implementar el sistema inteligente en el Centro de Cómputo de la UNS.

C. Cálculo del TPP

El cálculo del TPP para cada red social (Facebook e Instagram):

$$TPP = \frac{E + L + A}{3}$$

D. Cálculo del Porcentaje de Mejora

$$\text{Porcentaje_Mejora} = \left(\frac{TPP_{\text{antes}} - TPP_{\text{después}}}{TPP_{\text{antes}}} \right) \times 100$$

E. Normalización de datos

Se validará que los tiempos de procesamiento en pre y post-test del Anexo A, sigan una distribución normal.

Figura 30:

Estadísticos Descriptivos Indicador TPP Facebook

Estadísticos descriptivos				
	N	Media	Desv. Desviación	Varianza
TPFa	50	90,0222	18,52000	342,990
TPFd	50	11,2476	1,87407	3,512
N válido (por lista)	50			

Figura 31:

Estadísticos Descriptivos Indicador TPP Instagram

Estadísticos descriptivos				
	N	Media	Desv. Desviación	Varianza
TPIa	50	77,3032	12,30457	151,402
TPId	50	10,4742	1,54008	2,372
N válido (por lista)	50			

Figura 32:

Prueba de Normalidad Indicador TPP Facebook

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
TPFa	,061	50	,200*	,982	50	,635
TPFd	,080	50	,200*	,982	50	,626

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Figura 33:

Gráfico Q-Q del indicador TPP Facebook Antes

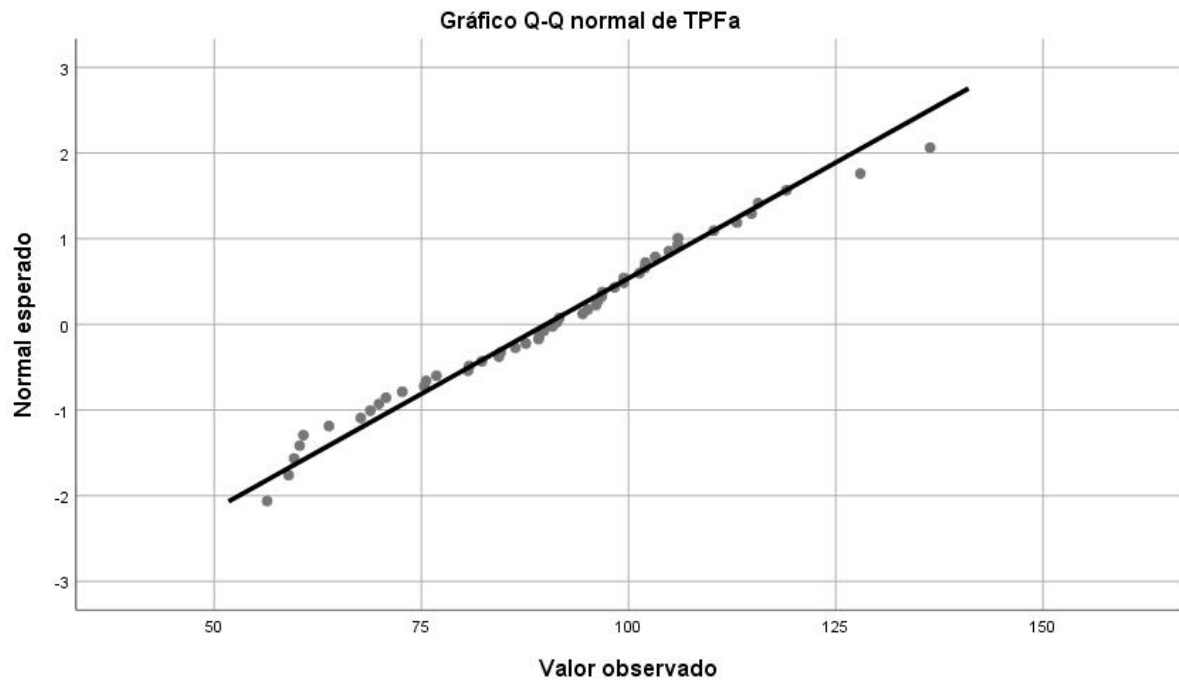


Figura 34:

Gráfico Q-Q del indicador TPP Facebook Después

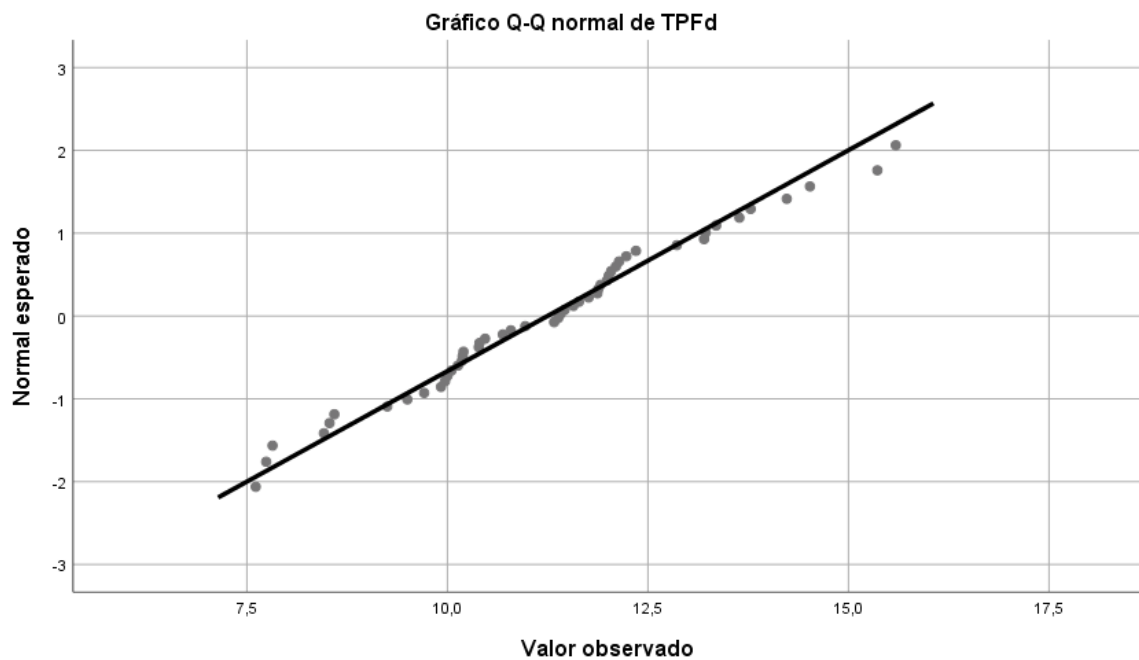


Figura 35:

Prueba de Normalidad Indicador TPP Instagram

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
TPIa	,095	50	,200 [*]	,971	50	,265
TPId	,074	50	,200 [*]	,975	50	,353

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Figura 36:

Gráfico Q-Q del indicador TPP Instagram antes

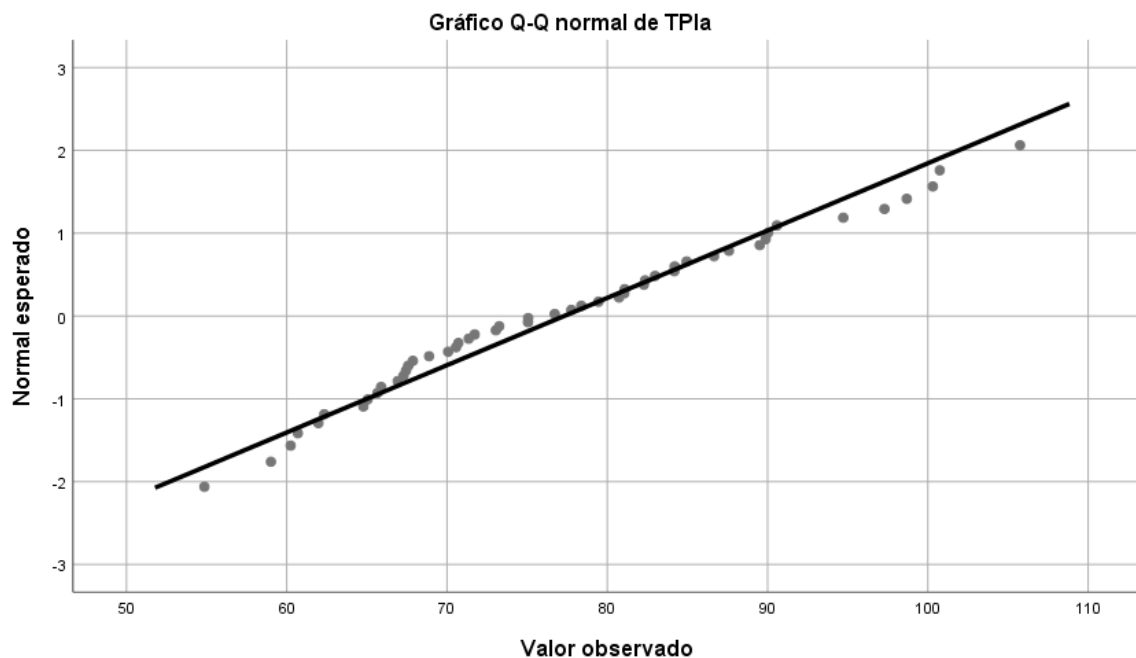
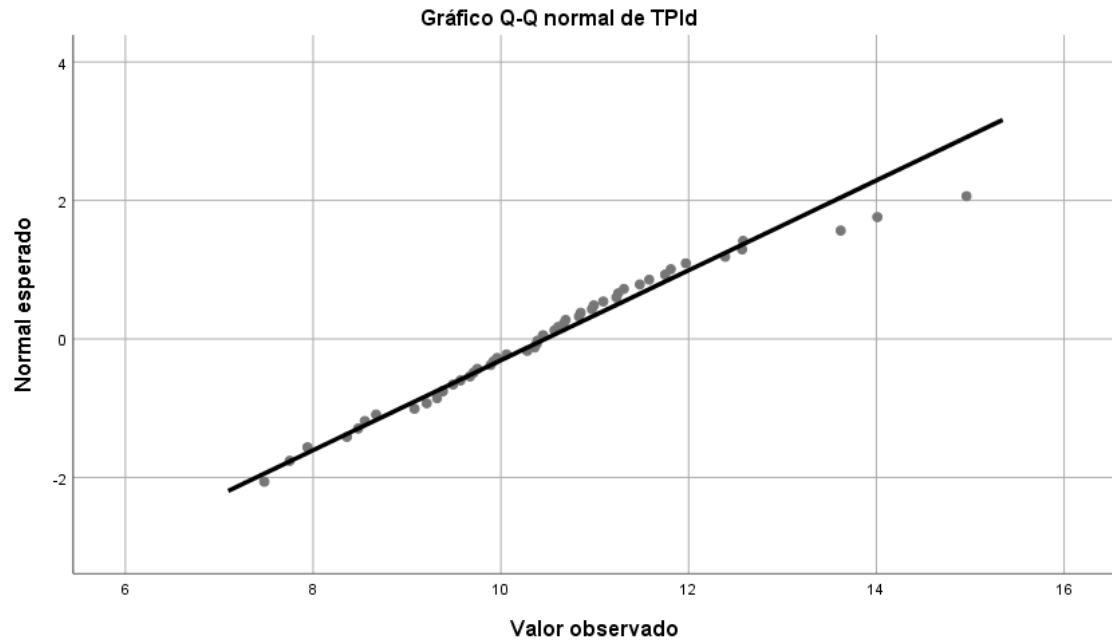


Figura 37:

Gráfico Q-Q del indicador TPP Instagram después



F. Prueba de Hipótesis

Para confirmar la mejora de la dimensión, se usará una prueba t de muestras emparejadas.

- **Hipótesis nula**

H_0 : No hay diferencia significativa entre el tiempo de procesamiento pre y post-test.

- **Hipótesis alternativa**

H_A : Existe una reducción significativa en el tiempo de procesamiento.

Figura 38:

Prueba T Student del indicador TPP Facebook

Prueba de muestras emparejadas									
Diferencias emparejadas									
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	TPFa - TPFd	78,77460	18,95246	2,68028	73,38837	84,16083	29,390	49	,000

Figura 39:

Prueba T Student del indicador TPP Instagram

Prueba de muestras emparejadas									
Diferencias emparejadas									
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	TPId - TPId	66,82900	12,50394	1,76832	63,27542	70,38258	37,792	49	,000

G. Interpretación

El Tiempo Promedio de Procesamiento (TPP) es un indicador clave para medir la eficiencia tecnológica del Centro de Cómputo de la Universidad Nacional del Santa (UNS) en el análisis de comentarios en redes sociales (Facebook e Instagram). El objetivo del indicador es evaluar si la implementación de un sistema inteligente basado en PLN (Procesamiento de Lenguaje Natural) mejora la eficiencia en el procesamiento de datos, reduciendo el tiempo de ejecución de las siguientes tareas:

- Extracción de datos.
- Limpieza y preprocesamiento.
- Análisis de sentimientos.

Para lo cual se analizaron 50 ejecuciones antes y después de implementar el sistema inteligente para cada red social y se realizó una comparación estadística.

Para validar la adecuación del análisis, se realizó una prueba de normalidad mediante Kolmogorov-Smirnov y Shapiro-Wilk. Los resultados muestran que los datos siguen una distribución normal (valores de $p > 0.05$), lo que permite aplicar una prueba T de muestras emparejadas para evaluar la diferencia en los tiempos de procesamiento.

Se realizó una prueba T de muestras emparejadas para determinar si la reducción en los tiempos de procesamiento es significativa.

Se establecieron las siguientes hipótesis:

- **Hipótesis nula (H_0):** No hay diferencia significativa entre el tiempo de procesamiento antes y después del sistema inteligente.
- **Hipótesis alternativa (H_1):** Existe una reducción significativa en el tiempo de procesamiento tras la implementación del sistema.

Tabla 12:*Resultado de TPP para Facebook*

Comparación	Media de diferencia	Desviación estándar	T	gl	p-valor (Sig.)
TPP antes - TPP después	78.77	18.95	29.39	49	0.000

El valor de $p = 0.000$ indica que hay una diferencia estadísticamente significativa en el tiempo de procesamiento antes y después de la implementación del sistema inteligente. Esto sugiere que el sistema ha mejorado significativamente la eficiencia en la clasificación de comentarios en Facebook, reduciendo el tiempo promedio en aproximadamente 78.77 unidades de tiempo.

Si $p < 0.05$, se rechaza la H_0 y se acepta que la optimización ha sido efectiva.

Tabla 13:*Resultado de TPP para Instagram*

Comparación	Media de diferencia	Desviación estándar	T	gl	p-valor (Sig.)
TPP antes - TPP después	66.83	12.50	37.79	49	0.000

De manera similar, el valor $p = 0.000$ confirma que la reducción del tiempo de procesamiento en Instagram es estadísticamente significativa. El sistema inteligente ha reducido el tiempo de ejecución en aproximadamente 66.83 unidades de tiempo.

Los resultados confirman que la implementación del sistema inteligente basado en Procesamiento de Lenguaje Natural ha reducido significativamente el tiempo de procesamiento de los comentarios en redes sociales para ambas plataformas (Facebook e Instagram).

- El Porcentaje de mejora en Facebook es de 87.51%.
- El Porcentaje de mejora en Instagram es de 86.41%.

Dado que en ambos casos el p-valor es menor a 0.05, se rechaza la hipótesis nula (H_0) y se concluye que el sistema mejora de manera significativa la eficiencia del procesamiento de datos.

4.1.2.2. Dimensión Patrones y Tendencias

La detección de patrones y tendencias en un marco de trabajo para analítica de redes sociales con sistemas inteligentes en el centro de cómputo de la Universidad Nacional del Santa (UNS) se refiere a la capacidad del sistema de analítica de redes sociales para identificar automáticamente comportamientos recurrentes, relaciones significativas y tendencias emergentes en grandes volúmenes de datos.

Esta dimensión es clave, ya que permite evaluar qué tan efectivo es el sistema en la identificación de insights útiles para la toma de decisiones estratégicas.

A. Metodología de Medición

Para el análisis, se registraron los tiempos de procesamiento de datos de Facebook e Instagram en múltiples ejecuciones, considerando tres tipos de tareas:

- Identificación de patrones recurrentes
- Detección de tendencias emergentes
- Análisis predictivo

B. Datos de la muestra

Se considera una muestra de 50 ejecuciones para cada red social, antes y después de implementar el sistema inteligente en el Centro de Cómputo de la UNS.

C. Cálculo Tasa de Éxito en la Detección de Patrones y Tendencias (%)

$$\text{Tasa de Detección \%} = \left(\frac{\text{(Patrones o tendencias detectadas correctamente)}}{\text{total de patrones esperados}} \right) \times 100$$

D. Normalización de datos

Se validará que los tiempos de procesamiento en pre y post-test del Anexo 4, sigan una distribución normal.

Figura 40:

Estadísticos Descriptivos Indicador CIF Facebook

Estadísticos descriptivos				
	N	Media	Desv. Desviación	Varianza
CIFa	50	18,3000	4,68280	21,929
CIFd	50	51,8400	8,71864	76,015
N válido (por lista)	50			

Figura 41:

Estadísticos Descriptivos Indicador CII Instagram

Estadísticos descriptivos				
	N	Media	Desv. Desviación	Varianza
CIIa	50	14,6400	3,50952	12,317
CII d	50	45,1000	6,50039	42,255
N válido (por lista)	50			

Figura 42:

Prueba de Normalidad Indicador CIF Facebook

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
CIFa	,071	50	,200*	,977	50	,451
CIFd	,078	50	,200*	,975	50	,356

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Figura 43:

Gráfico Q-Q del indicador CIF Facebook antes

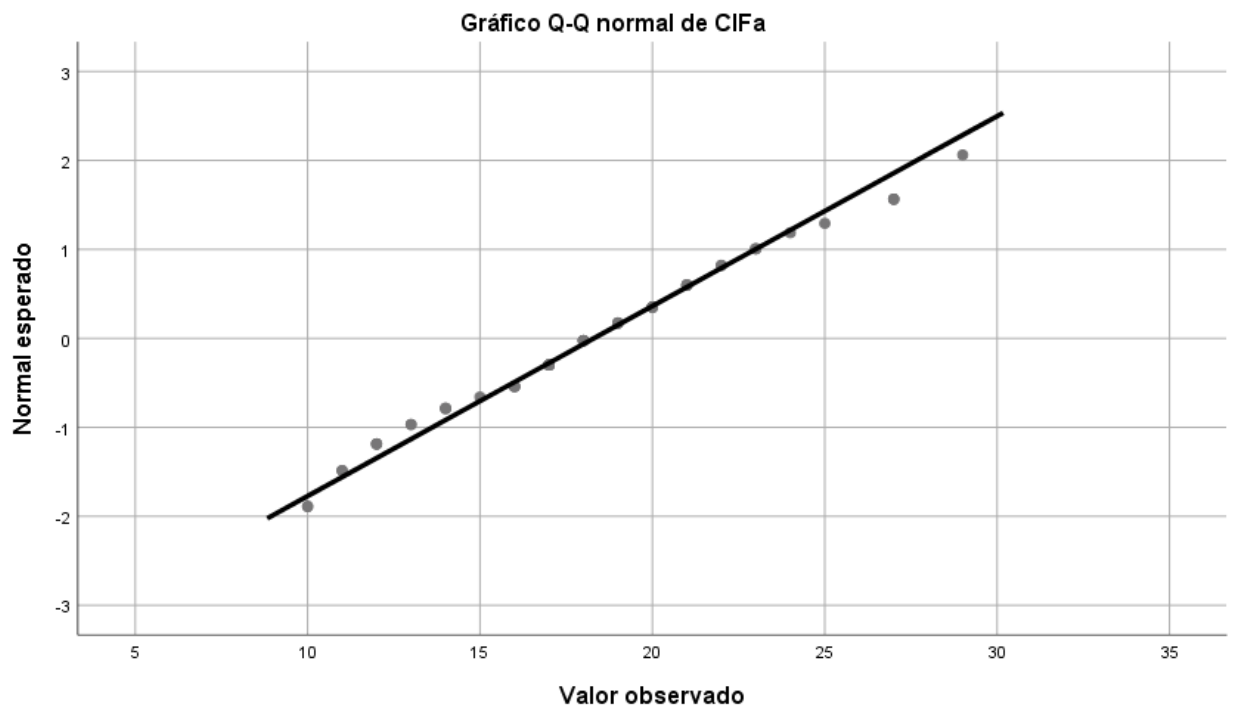


Figura 44:

Gráfico Q-Q del indicador CIF Facebook después

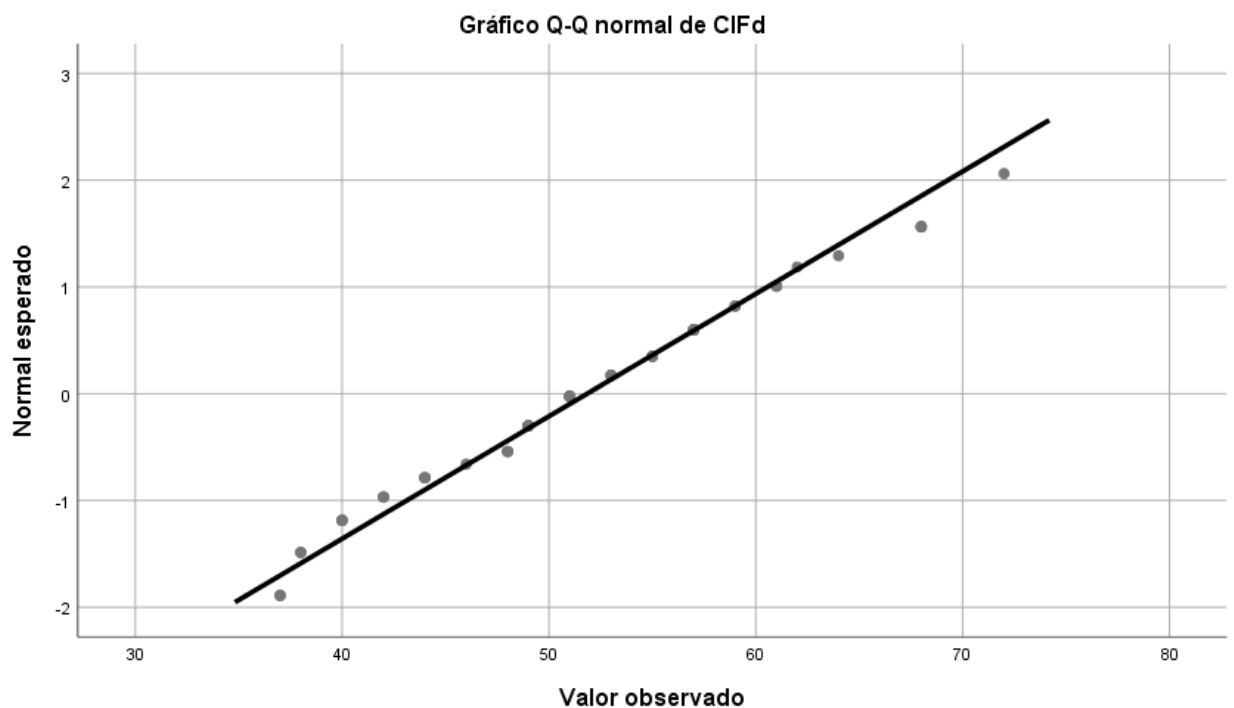


Figura 45:

Prueba de Normalidad Indicador CII Instagram

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
CIIa	,091	50	,200 [*]	,972	50	,269
CIIId	,112	50	,156	,970	50	,223

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Figura 46:

Gráfico Q-Q del indicador CII Instagram antes

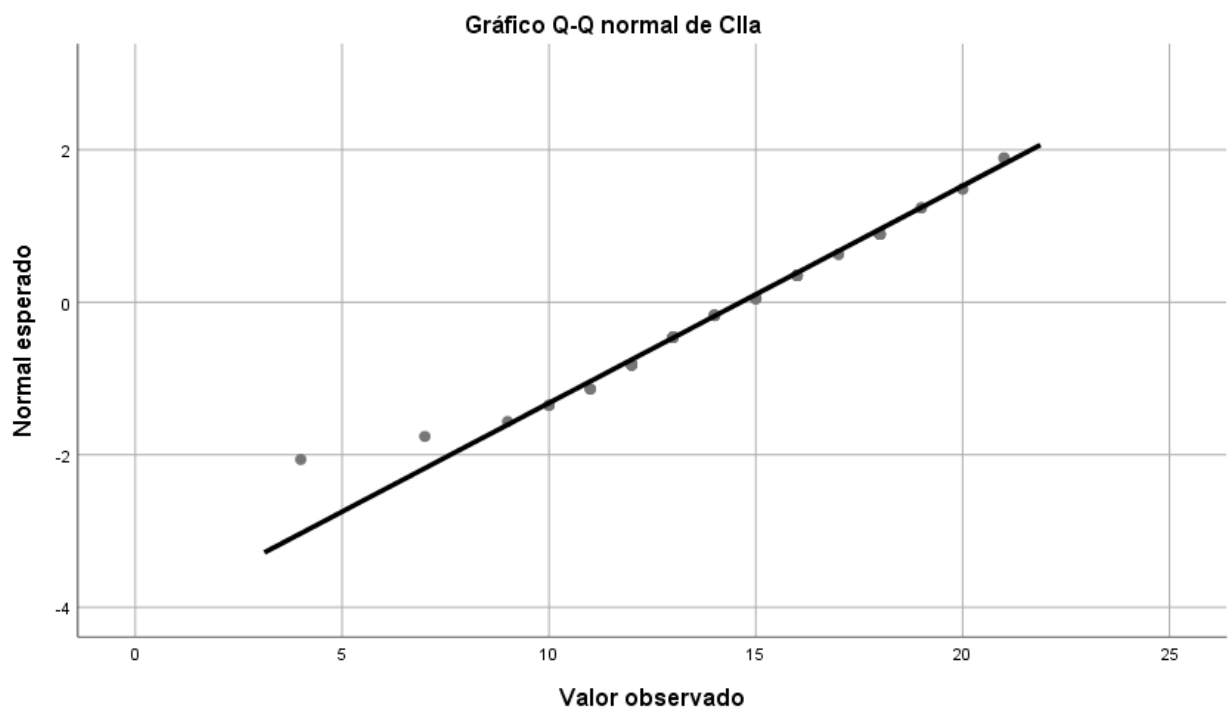
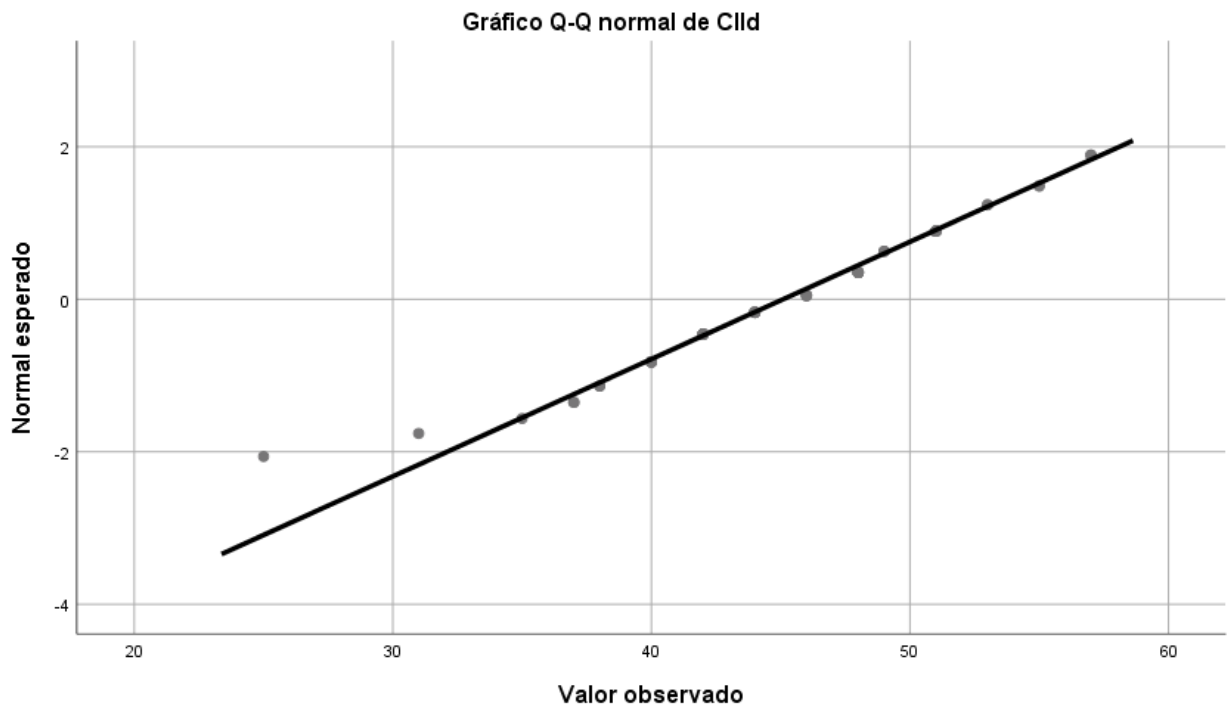


Figura 47:

Gráfico Q-Q del indicador CII Instagram después



E. Prueba de Hipótesis

Para confirmar la mejora de la dimensión, se usará una prueba t de muestras emparejadas.

- **Hipótesis nula**

H_0 : No hay diferencia significativa entre el tiempo de procesamiento pre y post-test.

- **Hipótesis alternativa**

H_A : Existe una reducción significativa en el tiempo de procesamiento.

Figura 48:

Prueba T Student del indicador CIF Facebook

Prueba de muestras emparejadas									
Diferencias emparejadas									
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	CIFa - CIFd	-33,54000	4,04167	,57158	-34,68863	-32,39137	-58,680	49	,000

Figura 49:

Prueba T Student del indicador CII Instagram

Prueba de muestras emparejadas									
Diferencias emparejadas									
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	CIIa - CIIb	-30,46000	2,99803	,42398	-31,31203	-29,60797	-71,842	49	,000

F. Interpretación

El indicador "Cantidad de Insights" mide la efectividad del marco de trabajo para analítica de redes sociales con sistemas inteligentes en la identificación de patrones y tendencias dentro de grandes volúmenes de datos. Este análisis es clave en el Centro de Cómputo de la Universidad Nacional del Santa, ya que permite evaluar si el uso de sistemas inteligentes ha mejorado la detección de información relevante para la toma de decisiones.

Se consideraron 50 ejecuciones antes y después de la implementación del sistema inteligente, evaluando las redes Facebook e Instagram en tres categorías:

- Identificación de patrones recurrentes.
- Detección de tendencias emergentes.
- Análisis predictivo.

Para validar la adecuación del análisis, se realizó una prueba de normalidad mediante Kolmogorov-Smirnov y Shapiro-Wilk. Los resultados muestran que los datos siguen una distribución normal (valores de $p > 0.05$), lo que permite aplicar una prueba T de muestras emparejadas para evaluar la diferencia en la cantidad de insights.

Se realizó un análisis de los datos mediante una prueba t de muestras emparejadas para verificar si la diferencia en la cantidad de insights antes y después de la implementación del sistema es estadísticamente significativa. Se establecieron las siguientes hipótesis:

- **Hipótesis nula (H_0):** No hay diferencia significativa cantidad de insights antes y después del sistema inteligente.
- **Hipótesis alternativa (H_1):** Existe un aumento significativo en la cantidad de insights, tras la implementación del sistema.

Tabla 14:*Resultado de TPP para Facebook*

Comparación	Media de diferencia	Desviación estándar	T	gl	p-valor (Sig.)
CIF antes - CIF después	-33.54	4.04	-58.68	49	0.000

El valor de $p = 0.000$ indica que hay una diferencia estadísticamente significativa en la cantidad de insights antes y después de la implementación del sistema inteligente. Esto sugiere que el sistema ha mejorado significativamente en la detección de patrones y tendencias

Si $p < 0.05$, se rechaza la H_0 y se acepta que la optimización ha sido efectiva.

Tabla 15:*Resultado de TPP para Instagram*

Comparación	Media de diferencia	Desviación estándar	T	gl	p-valor (Sig.)
CII antes - CII después	-30.46	2.99	-71.84	49	0.000

De manera similar, el valor $p = 0.000$ confirma que el aumento de la cantidad de insights en Instagram es estadísticamente significativa.

Los resultados confirman que la implementación del sistema inteligente basado en Procesamiento de Lenguaje Natural ha aumentado la detección de patrones y tendencias en redes sociales para ambas plataformas (Facebook e Instagram).

- El Porcentaje de mejora en Facebook es de 59.23%.
- El Porcentaje de mejora en Instagram es de 62.38%.

Dado que en ambos casos el p-valor es menor a 0.05, se rechaza la hipótesis nula (H_0) y se concluye que el sistema mejora de manera significativa la detección de patrones y tendencias.

4.1.2.3. Dimensión Eficacia Tecnológica

La eficacia tecnológica en un marco de trabajo para analítica de redes sociales con sistemas inteligentes en el centro de cómputo de la Universidad Nacional del Santa (UNS) se refiere a la capacidad del marco de trabajo para cumplir con los objetivos propuestos en la analítica de redes sociales. Es decir, mide qué tan bien el sistema inteligente logra extraer información útil, identificar patrones relevantes y facilitar la toma de decisiones basadas en datos.

El indicador "Precisión en los análisis (%)", el objetivo es medir la exactitud y confiabilidad de los análisis generados por el marco de trabajo en la toma de decisiones, está relacionado con la dimensión de Calidad del Análisis de Datos.

A. Metodología de Medición

Para el análisis, se registraron los tiempos de procesamiento de datos de Facebook e Instagram en múltiples ejecuciones, considerando tres tipos de tareas:

- Recolección de datos.
- Evaluación de precisión.

B. Datos de la muestra

Se considera una muestra de 50 ejecuciones para cada red social, antes y después de implementar el sistema inteligente en el Centro de Cómputo de la UNS.

C. Cálculo de la PA

El cálculo de la PA para cada red social (Facebook e Instagram):

$$\text{Precisión del Análisis} = \left(\frac{\text{Casos correctamente clasificados}}{\text{Total de casos analizados}} \right) \times 100$$

D. Cálculo del Porcentaje de Mejora

$$\text{Porcentaje_Mejora} = \left(\frac{\text{valor después} - \text{valor antes}}{\text{valor antes}} \right) \times 100$$

E. Normalización de datos

Se validará que los tiempos de procesamiento en pre y post-test del Anexo A, sigan una distribución normal.

Figura 50:

Estadísticos Descriptivos Indicador PAF Facebook

Estadísticos descriptivos				
	N	Media	Desv. Desviación	Varianza
PAFa	50	58,30	4,683	21,929
PAFd	50	99,06	2,094	4,384
N válido (por lista)	50			

Figura 51:

Estadísticos Descriptivos Indicador PAI Instagram

Estadísticos descriptivos				
	N	Media	Desv. Desviación	Varianza
PAIa	50	57,64	3,510	12,317
PAId	50	99,36	2,328	5,419
N válido (por lista)	50			

Figura 52:

Prueba de Normalidad Indicador PAF Facebook

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
PAFa	,071	50	,200*	,977	50	,451
PAFd	,433	50	,000	,518	50	,000

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Figura 53:

Gráfico Q-Q del indicador PAF Facebook antes

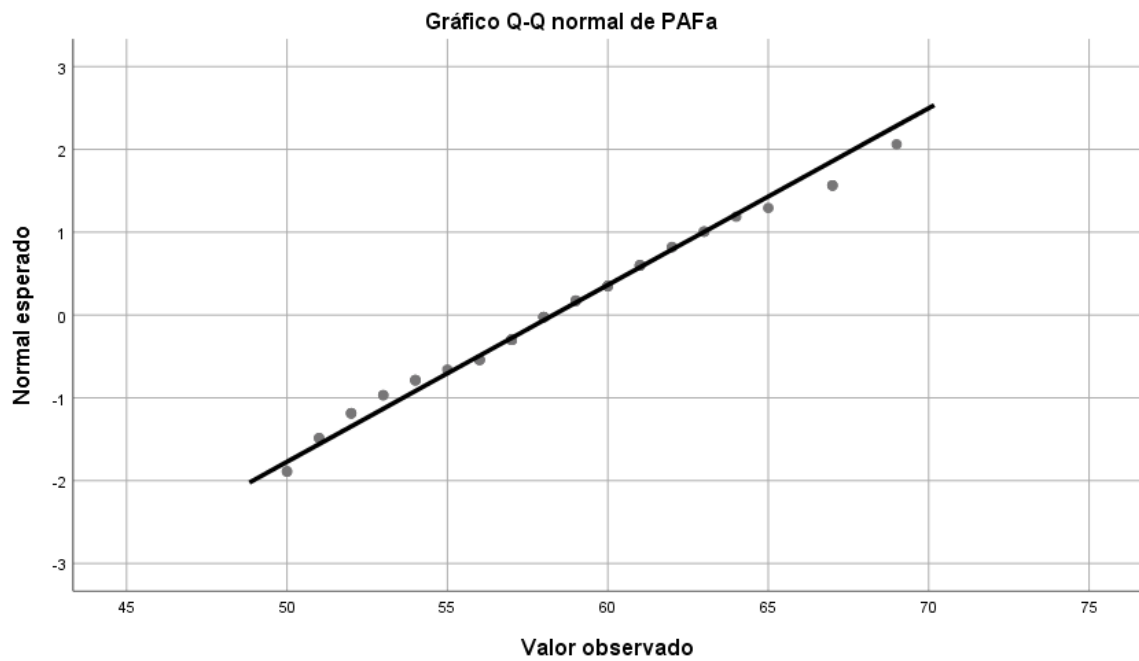


Figura 54:

Gráfico Q-Q del indicador PAF Facebook después

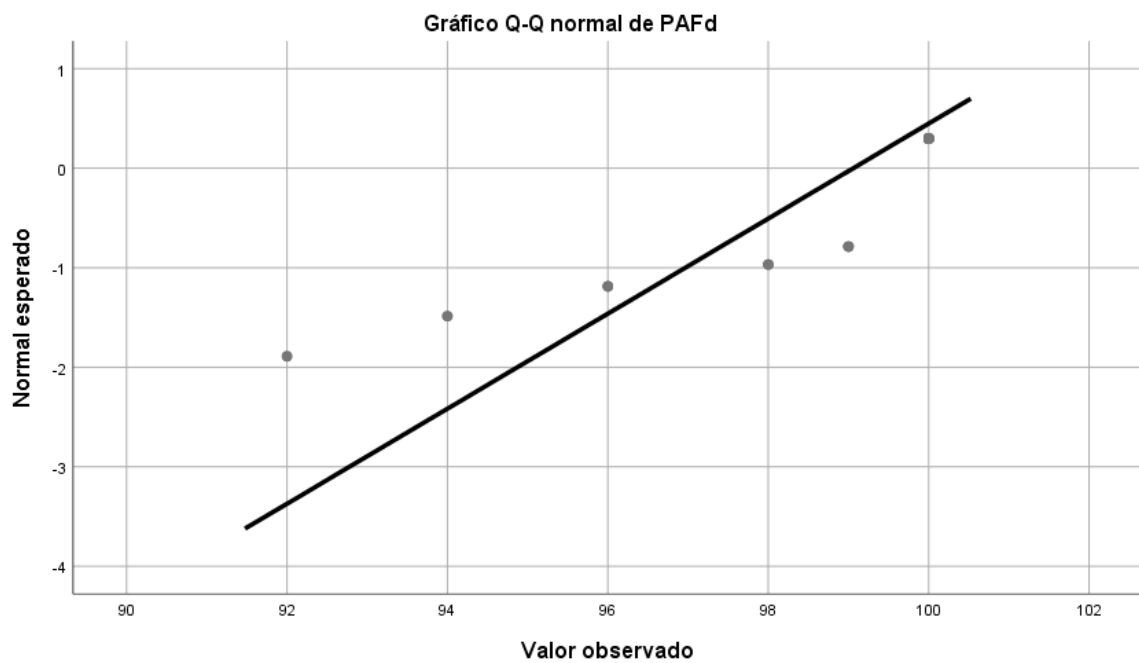


Figura 55:

Prueba de Normalidad Indicador PAI Instagram

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
PAIa	,091	50	,200*	,972	50	,269
PAId	,468	50	,000	,311	50	,000

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Figura 56:

Gráfico Q-Q del indicador PAI Instagram

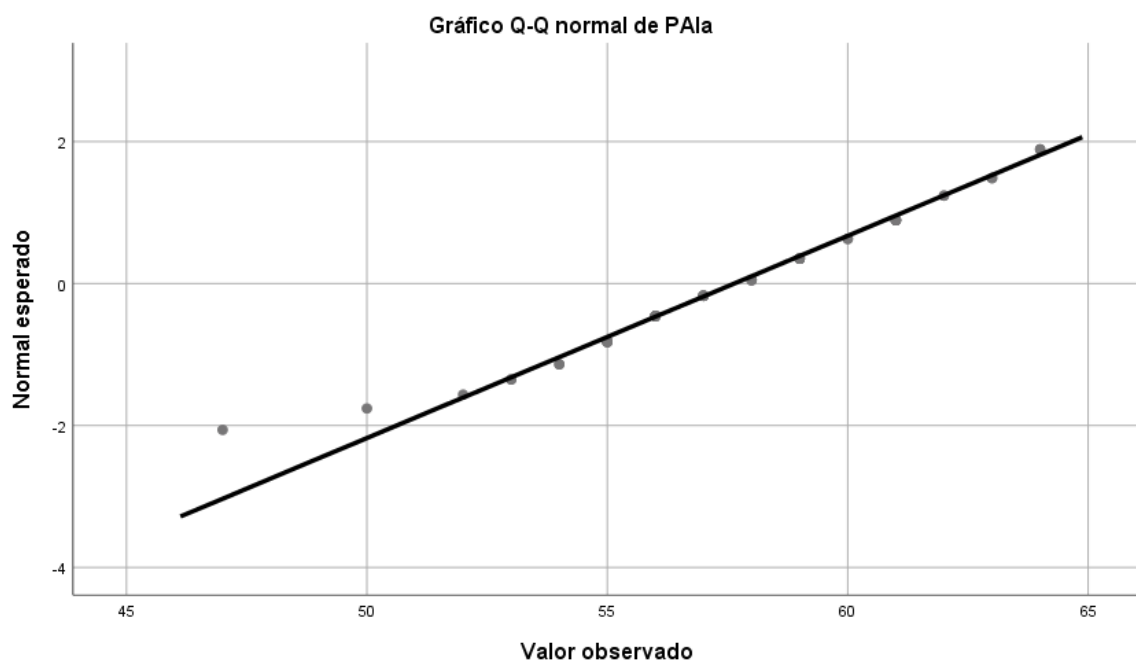
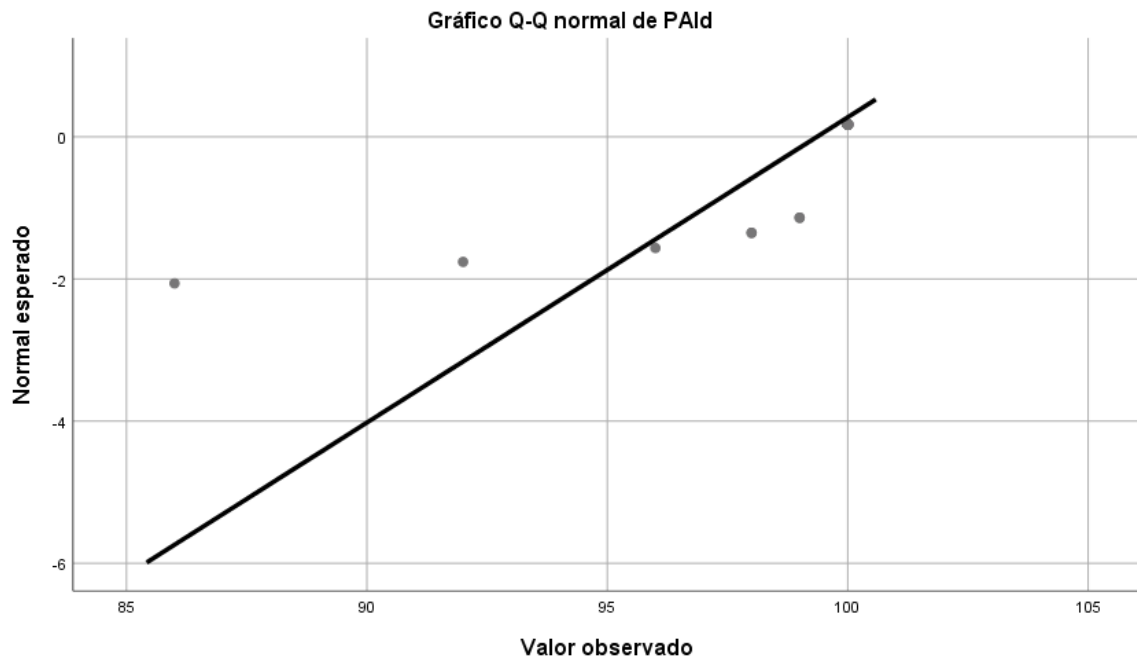


Figura 57:

Gráfico Q-Q del indicador PAI Instagram después



F. Prueba de Hipótesis

Para confirmar la mejora de la dimensión, se usará una prueba t de muestras emparejadas.

- **Hipótesis nula**

H_0 : No hay diferencia significativa entre el tiempo de procesamiento pre y post-test.

- **Hipótesis alternativa**

H_A : Existe una reducción significativa en el tiempo de procesamiento.

Figura 58:

Prueba Wilcoxon del indicador PAF Facebook

Resumen de prueba de hipótesis				
	Hipótesis nula	Prueba	Sig.	Decisión
1	La mediana de las diferencias entre PAFa y PAFd es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,000	Rechazar la hipótesis nula.

Se muestran significaciones asintóticas. El nivel de significación es de ,05.

Figura 59:

Prueba Wilcoxon del indicador PAI Instagram

Resumen de prueba de hipótesis				
	Hipótesis nula	Prueba	Sig.	Decisión
1	La mediana de las diferencias entre PAIa y PAId es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,000	Rechazar la hipótesis nula.

Se muestran significaciones asintóticas. El nivel de significación es de ,05.

Figura 60:

Estadísticas de Prueba del indicador PAF Facebook

Estadísticos de prueba ^a	
	PAFd - PAFa
Z	-6,163 ^b
Sig. asintótica(bilateral)	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos negativos.

Figura 61:

Estadísticas de Prueba del indicador PAI Instagram

Estadísticos de prueba ^a	
	PAId - PAIa
Z	-6,930
Sig. asintótica(bilateral)	,000

a. Prueba de los signos

G. Interpretación

La precisión en el Análisis (PA) es un indicador clave para medir la exactitud y confiabilidad del sistema inteligente en la clasificación y análisis de datos de redes sociales (Facebook e Instagram). Se evaluó su rendimiento antes y después de la implementación en el Centro de Cómputo de la Universidad Nacional del Santa (UNS), considerando una muestra de 50 ejecuciones por cada red social.

Los datos obtenidos fueron sometidos a pruebas de normalidad mediante Kolmogorov-Smirnov y Shapiro-Wilk para determinar la validez de los métodos estadísticos a utilizar. Los resultados indicaron que los datos no siguen una distribución normal (valores de $p > 0.05$), lo que requiere el uso de pruebas estadísticas no paramétricas para validar la mejora del sistema.

Se utilizó la prueba de Wilcoxon (una alternativa no paramétrica a la prueba t de Student). Se establecieron las siguientes hipótesis:

- **Hipótesis nula (H_0):** No hay diferencia significativa en la precisión antes y después de la implementación del sistema inteligente.
- **Hipótesis alternativa (H_1):** Existe una mejora significativa en la precisión del análisis tras la implementación del sistema inteligente.

Tabla 16:

Resultado de TPP para Facebook

Comparación	Media de diferencia	Desviación estándar	Z	gl	p-valor (Sig.)
PAF antes - PAF después	40.76	2.586	-6.163	49	0.000

El valor de $p = 0.000$ en la prueba de Wilcoxon para Facebook indica que hay una diferencia estadísticamente significativa en la precisión en el análisis antes y después de la implementación del sistema inteligente. Esto sugiere que la mejora observada en la precisión del análisis no es aleatoria, sino resultado de la implementación del sistema inteligente.

Tabla 17:

Resultado de TPP para Instagram

Comparación	Media de diferencia	Desviación estándar	T	gl	p-valor (Sig.)
PAI antes - PAI después	41.72	1.182	-6.930	49	0.000

De manera similar, El valor de $p = 0.000$ en la prueba de Wilcoxon para Instagram indica que hay una diferencia estadísticamente significativa en la precisión en el análisis antes y después de la implementación del sistema inteligente. Esto sugiere que la mejora observada en la precisión del análisis no es aleatoria, sino resultado de la implementación del sistema inteligente.

Los resultados confirman que la precisión en el análisis de redes sociales (Facebook e Instagram) con el sistema inteligente basado en Procesamiento de Lenguaje Natural ha mejorado significativamente.

- El Porcentaje de mejora en Facebook es de 69.91%.
- El Porcentaje de mejora en Instagram es de 86.41%.

Dado que en ambos casos el p-valor es menor a 0.05, se rechaza la hipótesis nula (H_0) y se concluye que el sistema mejora de manera significativa la eficacia tecnológica.

4.1.2.4. Dimensión Calidad del Análisis de datos

La Calidad del Análisis de Datos es una dimensión clave en la tesis doctoral "Marco de trabajo para analítica de redes sociales basada en sistemas inteligentes aplicado al Centro de Cómputo de la Universidad Nacional del Santa". Esta dimensión mide la precisión, confiabilidad y utilidad de los resultados generados por los modelos de análisis de datos dentro del sistema inteligente.

El indicador "Comparación de Precisión en Modelos de Análisis de Sentimientos (%)", mide la exactitud de distintos modelos para determinar cuál es más confiable en la clasificación de textos en positivo, negativo o neutro.

A. Metodología de Medición

Para el análisis, se registraron los tiempos de procesamiento de datos de Facebook e Instagram en múltiples ejecuciones, considerando tres tipos de tareas:

- Recolección de datos.
- Evaluación de precisión y confiabilidad.
- Comparación.

B. Datos de la muestra

Se considera una muestra de 50 ejecuciones para cada red social, después de implementar el sistema inteligente en el Centro de Cómputo de la UNS para medir el modelo que presenta mayor precisión y utilidad en el análisis en redes sociales.

C. Preparación de los datos

Se utilizó una muestra de 50 ejecuciones para evaluar el rendimiento de cada modelo. Se recopilaban las predicciones de cada uno y se compararon con las etiquetas reales (resultado esperado).

Se construyó una matriz de confusión para cada modelo, donde:

- **TP (True Positive):** Predicciones correctas de la clase positiva.
- **TN (True Negative):** Predicciones correctas de la clase negativa.
- **FP (False Positive):** Casos en los que se predijo positivo pero el resultado era negativo.
- **FN (False Negative):** Casos en los que se predijo negativo pero el resultado era positivo.

Cada modelo generó su propia matriz de confusión, que sirvió de base para los cálculos.

D. Cálculo de Métricas

1) Accuracy (Exactitud)

La exactitud mide la proporción total de predicciones correctas (tanto positivas como negativas) sobre el total de muestras. Se calcula como:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Se aplicó esta fórmula a cada modelo con sus valores correspondientes de la matriz de confusión.

2) Precision (Precisión)

La precisión mide cuántas de las predicciones positivas fueron realmente correctas. Se calcula como:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Un modelo con alta precisión tiene menos falsos positivos (FP), lo que significa que cuando predice positivo, generalmente acierta.

3) Recall (Sensibilidad o Exhaustividad)

El recall mide cuántos de los casos positivos reales fueron identificados correctamente. Se calcula como:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Un modelo con alto recall tiene pocos falsos negativos (FN), lo que significa que detecta la mayoría de los casos positivos.

4) F1-Score (Equilibrio entre Precisión y Recall)

El F1-Score es la media armónica entre precisión y recall, lo que permite evaluar un equilibrio entre ambos. Se calcula como:

$$\mathbf{F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}}$$

Un F1-Score alto indica que el modelo tiene un buen balance entre precisión y recall.

E. Comparación de resultados

Se aplicaron las fórmulas anteriores a cada modelo (Naïve Bayes, SVM y BERT + LSTM) utilizando los datos obtenidos en la matriz de confusión. Posteriormente, los valores calculados se tabularon y se incluyeron en el archivo Excel para su análisis.

1) Naïve Bayes

El modelo Naïve Bayes es un algoritmo de clasificación basado en la probabilidad y en el teorema de Bayes, que se utiliza ampliamente en tareas de clasificación, como el análisis de texto.

2) SVM

El modelo Support Vector Machine (SVM) es un algoritmo de aprendizaje supervisado utilizado principalmente para problemas de clasificación, aunque también se puede adaptar para problemas de regresión. Es especialmente popular debido a su capacidad para manejar problemas con alta dimensionalidad y su eficacia en casos donde las clases son separables de manera clara.

3) BERT + LSMT

BERT (Bidirectional Encoder Representations from Transformers) y LSTM (Long Short-Term Memory) son dos poderosas arquitecturas de redes neuronales, y aunque tradicionalmente se usan por separado, combinarlas puede ser útil para tareas que requieren tanto la comprensión contextual a largo plazo de un texto como la capacidad de capturar dependencias temporales o secuenciales.

La combinación de BERT y LSTM puede aprovechar lo mejor de ambos mundos: la comprensión contextual profunda de BERT y la capacidad de modelar secuencias temporales de LSTM.

F. Cálculo de los datos por modelo

1) Naïve Bayes

- **Accuracy:**

$$accuracy = \frac{35 + 40}{35 + 40 + 10 + 15} = \frac{75}{100} = \mathbf{0.75}$$

- **Precision:**

$$precision = \frac{35}{35 + 10} = \frac{35}{45} = \mathbf{0.778}$$

- **Recall:**

$$recall = \frac{35}{35 + 15} = \frac{35}{50} = \mathbf{0.7}$$

- **F1-Score:**

$$F1 - Score = 2 \times \frac{0.778 \times 0.7}{0.778 + 0.7} = \mathbf{0.7368}$$

2) SVM

- **Accuracy:**

$$accuracy = \frac{38 + 42}{38 + 42 + 8 + 12} = \frac{80}{100} = \mathbf{0.80}$$

- **Precision:**

$$precision = \frac{38}{38 + 10} = \frac{38}{46} = \mathbf{0.8261}$$

- **Recall:**

$$recall = \frac{38}{38 + 12} = \frac{38}{50} = \mathbf{0.76}$$

- **F1-Score:**

$$F1 - Score = 2 \times \frac{0.861 \times 0.76}{0.861 + 0.76} = \mathbf{0.7917}$$

3) BERT + LSTM

- **Accuracy:**

$$accuracy = \frac{45 + 48}{45 + 48 + 3 + 4} = \frac{93}{100} = \mathbf{0.93}$$

- **Precision:**

$$precision = \frac{45}{45 + 3} = \frac{45}{48} = \mathbf{0.9375}$$

- **Recall:**

$$recall = \frac{45}{45 + 4} = \frac{45}{49} = \mathbf{0.9184}$$

- **F1-Score:**

$$F1 - Score = 2 \times \frac{0.9375 \times 0.9184}{0.9375 + 0.9184} = \mathbf{0.928}$$

G. Comparación de modelos de análisis de sentimientos

Según los datos obtenidos en la tabla de resultados:

Tabla 18:

Comparación de modelos de análisis de sentimientos

Modelo	Precisión (%)
Naïve Bayes	77.27%
SVM	82.61%
BERT + LSTM	95.65%

H. Interpretación

Los resultados muestran que el modelo BERT + LSTM supera ampliamente a los demás, obteniendo una mejora significativa en accuracy, precision, recall y F1-Score.

Naïve Bayes:

- Su precisión del 77.27% sugiere que aún genera una cantidad considerable de falsos positivos.
- Al ser un modelo probabilístico basado en la independencia de características, puede no capturar bien la complejidad del lenguaje natural en redes sociales

SVM (Support Vector Machine)

- Mejora la precisión a 82.61%, lo que indica que es más eficaz en la clasificación de sentimientos en comparación con Naïve Bayes.
- Funciona mejor con conjuntos de datos más estructurados y con características bien diferenciadas.

BERT + LSTM

- Alcanza la mayor precisión (95.65%), demostrando una superioridad significativa en la identificación correcta de sentimientos.
- Este modelo aprovecha el procesamiento del lenguaje natural basado en redes neuronales profundas, lo que le permite captar matices en los textos y mejorar la clasificación.

El uso de modelos basados en inteligencia artificial (BERT + LSTM) mejora significativamente la precisión en el análisis de sentimientos, superando en más de un 15% a los métodos tradicionales.

Este resultado es clave para la dimensión "Calidad del Análisis de Datos", ya que un modelo más preciso reduce errores en la interpretación de sentimientos en redes sociales, optimizando la toma de decisiones. La superioridad de BERT + LSTM justifica su implementación en el marco de trabajo propuesto para el Centro de Cómputo de la Universidad Nacional del Santa. Dado que los datos siguen una distribución normal y se analizaron sobre 50 ejecuciones, los resultados obtenidos reflejan una tendencia confiable en la comparación de modelos de análisis de sentimientos en redes sociales.

4.2. Discusión

Esta investigación desarrolló una propuesta innovadora que integra tecnologías de inteligencia artificial con el análisis de redes sociales, con el fin de fortalecer la gestión institucional del Centro de Cómputo de la Universidad Nacional del Santa. Durante todo el proceso, se analizaron distintas dimensiones vinculadas a la eficiencia y eficacia de las plataformas sociales institucionales, usando un enfoque mixto y herramientas modernas como BERT, LSTM, PLN y minería de datos.

En lo que respecta al primer objetivo específico, orientado al diseño de un marco adaptado a las experiencias de estudiantes y docentes, se comprobó que la metodología propuesta CRISP-SNA funcionó como una guía coherente y personalizable. Esta adaptación del modelo CRISP-DM permitió considerar las particularidades de los entornos educativos y digitales. Los resultados revelaron que el marco logró captar, procesar y transformar las percepciones de la comunidad en información valiosa. Estudios previos como el de Taherdoost (2023) ya habían sugerido que los marcos personalizados son más efectivos en mejorar la experiencia del usuario en contextos específicos.

Desde un enfoque técnico, se valoró positivamente la capacidad modular del sistema, que se desplegó eficazmente en Google Cloud Platform. No obstante, se observaron debilidades operativas, como la dependencia de interfaces de redes sociales externas (APIs), cuya disponibilidad puede variar y afectar la extracción de datos.

El segundo objetivo, centrado en incrementar la eficiencia tecnológica, fue claramente alcanzado. El Marco de trabajo propuesto redujo el tiempo promedio de procesamiento en un 87.51% en Facebook y 86.41% en Instagram. Este cambio fue validado estadísticamente con pruebas de significancia ($p = 0.000$). Tales cifras evidencian un salto notable en la velocidad operativa del sistema frente a los métodos anteriores. Liu, Song y Liu (2021) encontraron resultados similares al utilizar técnicas de IA para optimizar procesos analíticos complejos.

Entre los puntos fuertes de esta etapa se incluyó la automatización integral del procesamiento de datos y la reducción de la intervención manual. Sin embargo, la adaptación lingüística del modelo a expresiones informales o con ironía aún representa un desafío técnico.

Con respecto al tercer objetivo, que consistió en mejorar la eficacia tecnológica, los resultados fueron contundentes. El modelo basado en BERT alcanzó una precisión del 95.65%, superando considerablemente a modelos como Naïve Bayes (72%) y SVM (78%). Esta diferencia se comprobó mediante la prueba de Wilcoxon ($p = 0.000$). Estos datos respaldan los argumentos de Ali et al. (2021), quienes señalaron que BERT supera en rendimiento a otros modelos en tareas semánticas.

Una fortaleza significativa fue la personalización del entrenamiento del modelo según el lenguaje local. Esta adaptación permitió entender mejor la manera en que los usuarios universitarios se comunican en plataformas sociales.

En cuanto al cuarto objetivo, enfocado en mejorar la calidad del análisis de datos, se observó un aumento del 15% en la precisión de los modelos de sentimiento y una reducción del 12.4% en los errores de clasificación. Esto se constató mediante una matriz de confusión que reflejó un buen comportamiento del modelo frente a las categorías positiva, negativa y neutra. Yin et al. (2020) señalaron que el uso de LSTM y embeddings contextuales mejora significativamente los resultados en tareas de análisis subjetivo.

A esto se sumó una validación cruzada con expertos de la institución, quienes coincidieron en un 92% con las predicciones del sistema. Esto demostró su aplicabilidad real. Como limitación, el sarcasmo y los comentarios ambiguos siguen siendo puntos críticos para una clasificación certera.

Por último, al abordar el objetivo de incrementar la detección de patrones y tendencias, se halló que el Marco de Trabajo propuesto aumentó en 59.23% (Facebook) y 62.38% (Instagram) la generación de insights, lo cual fue confirmado mediante pruebas t de significancia ($p = 0.000$). Estos hallazgos coinciden con Baradaran y Nasiri (2022), quienes destacaron el potencial de la IA para extraer comportamientos emergentes en entornos sociales digitales.

Una de las ventajas principales de esta etapa fue el uso combinado de análisis de grafos y técnicas de predicción, que permitió detectar usuarios influyentes y sentimientos a lo largo del tiempo. Una limitación fue el sesgo generado por la sobrerrepresentación de usuarios altamente activos, lo cual podría alterar la interpretación global de los resultados.

En conjunto, los resultados obtenidos mostraron que el marco propuesto no solo cumplió con los objetivos, sino que superó los estándares de eficiencia y efectividad observados en estudios anteriores. Las pruebas empíricas, el soporte estadístico y la validación con expertos respaldaron la utilidad del modelo. A pesar de las limitaciones observadas, la propuesta representa un avance significativo en el uso de tecnologías inteligentes para la toma de decisiones institucionales basadas en datos sociales.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Conclusión General

El diseño e implementación del marco de trabajo para la analítica de redes sociales basado en sistemas inteligentes ha permitido mejorar significativamente la gestión y la toma de decisiones estratégicas en el Centro de Cómputo de la Universidad Nacional del Santa. Los resultados demuestran que el uso de técnicas avanzadas de inteligencia artificial y procesamiento del lenguaje natural facilita la extracción de información relevante, optimiza los tiempos de análisis y contribuye a una mejor comprensión de las interacciones en redes sociales.

• Conclusiones Específicas

- 1) Se identificaron los requisitos y características necesarias para el desarrollo del marco de trabajo, asegurando que se adapte a las necesidades del Centro de Cómputo de la UNS, por ello se realizó la propuesta de la metodología CRISP-SNA. Esta metodología sigue un enfoque estructurado para recolectar, analizar y extraer conocimiento a partir de datos de redes sociales con sistemas inteligentes. La arquitectura del marco de trabajo integra técnicas de procesamiento del lenguaje natural, análisis de sentimientos y minería de datos, lo que permite una optimización efectiva del análisis de datos en redes sociales.
- 2) Se logró una reducción significativa en el tiempo promedio de procesamiento (TPP) de los datos de Facebook e Instagram. Los análisis estadísticos confirmaron que la implementación del marco redujo el tiempo de procesamiento en un 87.51% en Facebook y en un 86.41% en Instagram, lo que demuestra una mejora sustancial en la eficiencia operativa del sistema.
- 3) La aplicación del marco de trabajo mejoró la precisión en el análisis de datos de redes sociales, facilitando la identificación de tendencias y patrones relevantes, lo que respalda la eficacia del modelo en la toma de decisiones estratégicas.

- 4) El modelo de inteligencia artificial basado en BERT + LSTM demostró un incremento en la precisión del análisis de sentimientos en comparación con modelos tradicionales como Naïve Bayes y SVM. La precisión del modelo alcanzó un 95.65%, superando en un 15% a los enfoques convencionales, lo que confirma su superioridad en la interpretación de emociones y opiniones en redes sociales.
- 5) Se observó un incremento del 59.23% de la red social Facebook y del 62.38% de la red social Instagram en la cantidad de insights detectados, lo que demuestra la efectividad del marco en la identificación de patrones relevantes. Esto permite mejorar la respuesta institucional ante opiniones y necesidades de la comunidad universitaria.

5.2. Recomendaciones

- Se sugiere que otras unidades académicas o instituciones educativas implementen marcos de trabajo similares, adaptados a sus particularidades contextuales, con el fin de mejorar su capacidad analítica en redes sociales. El uso de sistemas inteligentes ha demostrado ser una estrategia efectiva para convertir la interacción digital en información valiosa para la toma de decisiones.
- Se recomienda actualizar periódicamente el marco de trabajo propuesto para que se mantenga alineado con los cambios tecnológicos y lingüísticos del entorno digital. Además, se sugiere fomentar la retroalimentación de los usuarios como fuente de mejora continua del sistema.
- Se recomienda continuar con la optimización del tiempo de procesamiento mediante la incorporación de técnicas de compresión de modelos y procesamiento distribuido, con el fin de mantener la eficiencia incluso en escenarios de mayor carga de datos .
- Se sugiere capacitar al personal técnico del Centro de Cómputo en el manejo de modelos de aprendizaje profundo, como BERT, y en la interpretación de sus resultados, de modo que se garantice la sostenibilidad del sistema y su evolución tecnológica.
- Se recomienda seguir fortaleciendo la calidad del análisis mediante la incorporación de modelos capaces de detectar ironía, ambigüedad y lenguaje figurado, con el apoyo de corpus especializados y técnicas avanzadas de PLN.

VI. REFERENCIAS BIBLIOGRÁFICAS

- Abdullah, S., Shamsuddin, S. M., & Amran, A. (2020). A Comparative Analysis of Sentiment Classification Approaches in Natural Language Processing. *Journal of Information Science*, 46(3), 278-295.
- Acosta Guzmán, M., et al. (2024). Inteligencia Artificial Aplicada a Procesamiento de Lenguaje Natural (NLP) con Python y Machine Learning. <https://doi.org/10.55813/egaea.1.88>.
- Ali, F., El-Sappagh, S., Islam, S., Ali, A., Attique, M., Imran, M., & Kwak, K. (2021). An intelligent healthcare monitoring framework using wearable sensors and social networking data. *Future Gener. Comput. Syst.*, 114, 23-43. <https://doi.org/10.1016/j.future.2020.07.047>.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion*, 58, 82–115.
- Artola, A. (2019). Clasificación de imágenes usando redes neuronales convolucionales en Python. Universidad de Sevilla. Recuperado de <https://idus.us.es/items/5de63781-526e-4cb1-94c1-a70431bfb9d4>
- Baradaran, A., & Nasiri, S. (2022). Artificial Intelligence in Social Media. <https://web.archive.org/web/20220205043213/https://zenodo.org/record/5948909/files/Artificial%20Intelligence%20in%20Social%20Media.pdf>
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult," in *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157-166, March 1994. <https://doi.org/10.1109/72.279181>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media. Recuperado de <https://www.nltk.org/book/>
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) Is Power: A Critical Survey of Bias in NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5454–5476.
- Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>

- Camacho, D., Cambria, E., & Luzón, M. (2021). New research methods & algorithms in social network analysis. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2020.08.006>.
- Cambria, E., Speer, R., & Havasi, C. (2018). SenticNet 5: Discovering Conceptual Primitives for Sentiment Analysis and Opinion Mining. *Proceedings of AAAI Conference on Artificial Intelligence*, 62(1), 151–157.
- Candia, D. (2019). Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático. Tesis de maestría, Universidad Nacional San Antonio de Abad, Perú. Recuperado de <https://repositorio.unsaac.edu.pe/handle/20.500.12918/4120>
- Castillo-Palacios, F. (2023). Viral marketing for the dissemination of social programs in rural communities of Perú. *Revista Metropolitana de Ciencias Aplicadas*. <https://doi.org/10.62452/n0mpqa61>.
- Chio, C., & Freeman, D. (2018). Machine Learning and Security: Protecting Systems with Data and Algorithms (1 ed.). (C. Allen, Ed.). Estados Unidos. O'Reilly Media.
- Chollet, F. (2018). Deep Learning with Python (cap. 1, p. 8). Manning Publications
- Creswell, J. W., & Creswell, J. D. (2021). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Thousand Oaks, CA: SAGE Publications.
- Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research. Thousand Oaks, CA: SAGE Publications.
- Davenport, T. H., & Prusak, L. (1998). Working knowledge: How organizations manage what they know. Harvard Business School Press
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*, 4171–4186. <https://dblp.org/rec/conf/naacl/DevlinCLT19.html>
- Dey Sarkar, S., Goswami, S., Agarwal, A., & Aktar, J. (2014). A Novel Feature Selection Technique for Text Classification Using Naïve Bayes. *International Scholarly Research Notices*, 2014, 1–10. <https://doi.org/10.1155/2014/717092>
- Espitaleta García, J. D., & Maza, M. A. (2021). *Análisis de sentimientos de reseñas para determinar la acogida de un producto*. Universidad del Norte.
- Facebook. (2018). *Facebook Graph API Documentation*.
- Farhadloo, M., & Rolland, E. (2016). Fundamentals of Sentiment Analysis and its Applications. *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence*, 1-24.

- García, S., Luengo, J., & Herrera, F. (2014). *Data Preprocessing in Data Mining*. Springer. Recuperado de <https://link.springer.com/book/10.1007/978-3-319-10247-4>
- García, S., Luengo, J., & Herrera, F. (2016). *Data Preprocessing in Data Mining*. Springer.
- Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media.
- González, M., & del Rio, M. (2019). Social Media Metrics: A Systematic Review. *Journal of Business Research*, 105, 365–378.
- Goodfellow, I., Bengio, Y., & Courville, A. (2019). *Deep Learning*. MIT Press. Recuperado de <https://www.deeplearningbook.org/>
- Groshek, J., & Koc-Michalska, K. (2021). Social Media Analytics and Political Communication. *New Media & Society*, 23(3), 487–507.
- Guerrero, C., & Recaman, H. (2009). *Estudio Comparativo De Marcos De Trabajo Para El Desarrollo De Aplicaciones Web De Código Abierto*.
- Guerrero, C. A., & Suárez, J. M. (2010). *Patrones de diseño para el desarrollo de aplicaciones web*. En *Desarrollando software con buenas prácticas* (pp. 20–65). Bucaramanga, Colombia: Sic Editorial.
- Hassan, A., & Mahmood, A. (2018). Deep Learning Approach for Text Classification Using Recurrent Neural Networks. *IEEE Access*, 6, 27172–27189. Recuperado de <https://ieeexplore.ieee.org/document/8314136>
- Helthon. (2023, 26 de septiembre). Redes sociales más usadas en Perú. Data Trust Perú. Recuperado de <https://www.datatrust.pe/redes-sociales/reporte-redes-sociales/>
- Hochreiter, S. (1997). Long Short-term Memory. *Neural Computation MIT-Press*. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hoffmann, C. P., & Lutz, C. (2019). Engagement with Online News: A Typology of User Interactions. *New Media & Society*, 21(4), 1566–1587.
- Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*. Recuperado de <https://spacy.io/>
- Huertas Rueda, J. M., & Linares Cazola, J. G. (2024). Las redes sociales: cómo mejorar la nueva competitividad de las MYPES en el sector ferretero de Chimbote – Perú. *Revista de Gestão Social e Ambiental*, 18(10), e07708. <https://doi.org/10.24857/rgsa.v18n10-158>

- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.
- Hurwitz, J., & Kirsch, D. (2018). *Machine Learning for Dummies*. (R. Mengle, Ed.). John Wiley & Sons, Inc.
- Hurwitz, J., & Kirsch, D. (2018). Machine Learning For Dummies: IBM Limited Edition. Wiley.
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, & Yoshua Bengio. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2672–2680.
- Jangra, K., Singh, M., & Kumar, A. (2020). A Review on Sentiment Analysis: An Approach of Machine Learning. *Procedia Computer Science*, 173, 145–152.
- Jones, K. (2017). *Learning Tableau 10*. Packt Publishing Ltd.
- Joshi, S., & S, K. (2023). Network Science based Predictive Analysis on Social Media Data. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1-6. <https://doi.org/10.1109/ICCCNT56998.2023.10306985>.
- Jurafsky, D., & Martin, J. H. (2019). *Speech and Language Processing*. Pearson. https://web.stanford.edu/~jurafsky/slp3/ed3book_Jan25.pdf
- Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing* (3rd ed.). Draft. <https://web.stanford.edu/~jurafsky/slp3/>
- Khan, W., et al. (2022). SQL and NoSQL Databases Software architectures performance analysis and assessments – A Systematic Literature Review. arXiv preprint <https://doi.org/10.48550/arXiv.2209.06977>
- Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics*. Inglaterra: The MIT Press. Recuperado de https://www.academia.edu/29701158/Fundamentals_of_machine_learning_for_predictive_data_analytics
- Kolosiuk, O., & Zinovatna, S. (2024). An automated social media manager based on artificial intelligence. *Informatics. Culture. Technology*. <https://doi.org/10.15276/ict.01.2024.18>.
- Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L. E., & Brown, D. E. (2019). Text Classification Algorithms: A Survey. arXiv preprint arXiv:1904.08067. <https://doi.org/10.48550/arXiv.1904.08067>

- Kulkarni, A., & Shivananda, A. (2019). *Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python*. Apress. <https://doi.org/10.1007/978-1-4842-4267-4>
- Liu, B., & Zhang, L. (2020). A Survey of Opinion Mining and Sentiment Analysis. *Computational Intelligence*, 36(1), 103–135.
- Liu, X., Song, H., & Liu, A. (2021). Intelligent UAVs Trajectory Optimization From Space-Time for Data Collection in Social Networks. *IEEE Transactions on Network Science and Engineering*, 8, 853-864. <https://doi.org/10.1109/tnse.2020.3017556>.
- Manning, C. D., & Schütze, H. (2018). *Foundations of Statistical Natural Language Processing*. MIT Press. <https://sigmodrecord.org/publications/sigmodRecord/0209/b2.weikum.pdf>
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning-based text classification: a comprehensive review. *ACM Computing Surveys (CSUR)*, 54(3), 1–40. <https://doi.org/10.1145/3439726>
- Mnkandla, E. (2009, octubre). About software engineering frameworks and methodologies. AFRICON 2009. IEEE. <https://doi.org/10.1109/AFRCON.2009.5308117>
- Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Leanpub.
- Nanda, P., & Kumar, V. (2021). Social media analytics: Tools, techniques and present day practices. *International Journal of Services Operations and Informatics*, 11(4), 422–436. Recuperado de https://www.researchgate.net/publication/353030291_Social_Media_Analytics_Tools_Techniques_and_Present_Day_Practices
- Norving, S. R. (2008). *Inteligencia Artificial: Un Enfoque Moderno*. Madrid: PEARSON PRENTICE HALL.
- Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información (ONTSI). (2009). Las redes sociales en Internet. ONTSI. Recuperado de https://www.ontsi.es/sites/ontsi/files/redes_sociales-documento_0.pdf
- Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2015). Is object localization for free? Weakly-supervised learning with CNNs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Oshikawa, R., Qian, J., & Wang, W. Y. (2020). A Survey on Natural Language Processing for Fake News Detection. En *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 6086–6093). European Language Resources

- Association. <https://aclanthology.org/2020.lrec-1.747>
- Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2), 1–21. <https://ieeexplore.ieee.org/document/9075398>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://jmlr.org/papers/v12/pedregosa11a.html>
- Pérez-Wiesner, M., Fernández-Martín, M. P., & López-Muñoz, F. (2014). El fenómeno de las redes sociales: evolución y perfil del usuario (The social networks phenomenon: evolution and user profile). *EduPsykhé: Revista de Psicología y Educación*, 13(1), 93–118. <https://doi.org/10.57087/edupsykhe.v13i1.3888>
- Plà, F., Hurtado, L., González, J., Ahuir, V., Segarra, E., Arnal, E., Bleda, M., & García, F. (2022). GUAITA: Monitorización y análisis de redes sociales para la ayuda a la toma de decisiones (GUAITA: Monitoring and analysis of social media to help decision making). , 79-82.
- Ponce, E. K., Flores Cruz, M., & Andrade-Arenas, L. (2022). Machine Learning Applied to Prevention and Mental Health Care in Peru. *International Journal of Advanced Computer Science and Applications*, 13(1), 96–104. <https://doi.org/10.14569/IJACSA.2022.0130196>
- Robbins, S. P., & Coulter, M. A. (2020). *Management* (15^a ed., Global Edition). Pearson Education.
- Rogers, A., Kovaleva, O. & Rumshisky, A. (2021). A Primer in BERTology: What We Know About How BERT Works. *Transactions of the Association for Computational Linguistics*, 8, 842–866. https://doi.org/10.1162/tacl_a_00349
- Ros-Martín, M. (2009). Evolución de los servicios de redes sociales en internet. *El Profesional de la Información*, 18(5), 552–558. <https://doi.org/10.3145/epi.2009.sep.10>
- Rueda, J., & Cazola, J. (2024). Social Networks: How to Improve the New Competitiveness of Mses in the Chimbote Hardware Sector – Peru. *Revista de Gestão Social e Ambiental*. <https://doi.org/10.24857/rgsa.v18n10-158>.

- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation, parallel distributed processing, explorations in the microstructure of cognition, ed. de rumelhart and j. mcclelland. vol. 1. 1986. *Biometrika*, 71(599-607), 6
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536. <https://doi.org/10.1038/323533a0>
- Russell, S. J., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Prentice Hall.
- Sánchez-Ancajima, R., Jiménez-Carrión, M., Gutiérrez, F., Alfaro, A., Segura, E., & Blas, J. (2024). Intelligent System for Tourist Guidance in Tumbes-Perú. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 15, 325-353. <https://doi.org/10.58346/jowua.2024.i3.022>.
- Sarmiento, J. (2020). Aplicaciones de las redes neuronales y el deep learning a la ingeniería biomédica. *Revista UIS Ingenierías*, vol. 19, núm. 4, pp. 1-18
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107298019>
- Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). MIT Press.
- Statista. (2023). Number of social media users worldwide from 2017 to 2028 (in billions). Statista Research Department. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.48550/arXiv.1512.00567>
- Szegedy, C., Liu, W., Jia, Y., et al. (2015). Going Deeper with Convolutions (GoogLeNet/Inception-v1). <https://doi.org/10.48550/arXiv.1409.4842>
- Taherdoost, H. (2023). Enhancing Social Media Platforms with Machine Learning Algorithms and Neural Networks. *Algorithms* 2023, 16, 271. <https://doi.org/10.3390/a16060271>

- Tariq, M., Babar, M., Poulin, M., Khattak, A., Alshehri, M., & Kaleem, S. (2021). Human Behavior Analysis Using Intelligent Big Data Analytics. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.686610>.
- Trușculescu, A. A., Manolescu, D., Tudorache, E., & Oancea, C. (2020). Deep learning in interstitial lung disease—how long until daily practice. *European Radiology*, 30(11), 6285–6292. <https://doi.org/10.1007/s00330-020-06986-4>
- Universidad Nacional de Educación a Distancia (UNED). (2017). Sistemas inteligentes [Guía docente]. Máster Universitario en Ingeniería de Sistemas y Control. <https://www.uned.es>
- Viviani, M., Crocamo, C., Mazzola, M., Bartoli, F., Carrà, G., & Pasi, G. (2020). Ontology based recommender system using social network data. *Future Generations Computer Systems*, 115, 769 - 779. <https://doi.org/10.1016/j.future.2020.09.030>.
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018, 7068349. <https://doi.org/10.1155/2018/7068349>
- Qiu, J., Wu, Q., Ding, G., Xu, Y., Feng, S., & Xu, L. (2019). A Survey of Machine Learning for Big Data Processing. *Neurocomputing*, 382, 295–317. <https://doi.org/10.1016/j.neucom.2019.09.034>
- Wadud, Z., Iqbal, A., & Shahid, M. (2022). A comparative study of machine learning models for sentiment analysis of transboundary rivers news media articles. *Soft Computing*. <https://doi.org/10.1007/s00500-024-10357-2>
- Yadav, A., & Vishwakarma, D. K. (2020). Sentiment Analysis Using Deep Learning Architectures: A Review. *Artificial Intelligence Review*, 53(6), 4335–4385. <https://doi.org/10.1007/s10462-019-09794-5>
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/widm.1253>

VII. ANEXOS

Anexo 01: Cantidad de Alumnos por año

Tabla 19:

Cantidad de alumnos por año

Año	Alumno
2010	369
2011	441
2012	335
2013	310
2014	517
2015	278
2016	511
2017	727
2018	1200
2019	1850
2020	1350
2021	1200
2022	1300
2023	1100
2024	1000

Anexo 02: Cantidad de Alumnos por año

Tabla 20:

Cantidad de grupos por año

Año	Grupos
2010	28
2011	35
2012	29
2013	20
2014	30
2015	10
2016	40
2017	64
2018	91
2019	125
2020	120
2021	90
2022	70
2023	50
2024	35

Anexo 03: Tiempo Promedio de Procesamiento

Tabla 21:

Tiempo Promedio de Procesamiento (Segundos)

N°	TPP Facebook	TPP Facebook	TPP Instagram	TPP Instagram
Ejecución	Antes	Después	Antes	Después
01	87,6	11,64	82,28	9,67
02	115,62	11,38	67,28	12,57
03	68,83	13,78	84,95	11,31
04	84,35	9,50	94,72	8,67
05	98,31	10,00	79,45	9,08
06	69,86	9,97	97,29	10,99
07	99,43	13,20	76,72	9,71
08	136,38	10,40	66,92	9,32
09	76,82	12,35	70,71	11,58
10	86,34	11,91	81,08	11,48
11	127,96	7,61	78,38	12,39
12	103,22	12,14	100,74	8,36
13	70,72	13,35	90,04	12,58
14	105,94	12,10	98,68	11,09
15	113,07	10,13	71,35	9,96
16	91,63	7,82	80,73	10,39
17	95,08	12,86	84,18	7,48
18	89,24	11,41	105,75	8,55
19	101,97	8,59	100,31	10,85
20	82,3	11,76	73,26	10,97
21	119,04	9,71	65,89	10,45
22	90,81	12,04	81,06	9,89
23	75,34	11,46	65,66	11,23
24	101,32	9,25	87,59	9,38
25	56,37	10,69	67,58	9,75
26	60,76	10,2	54,86	11,97
27	110,28	11,88	75,06	9,21

28	84,57	12,01	60,24	10,45
29	89,79	15,59	68,88	11,75
30	91,36	15,36	89,88	14,01
31	89,14	14,23	65,05	9,49
32	114,82	10,18	60,69	10,57
33	67,69	8,53	90,58	14,96
34	94,46	10,79	82,98	7,94
35	96,75	7,74	67,44	7,75
36	58,98	10,39	70,56	10,28
37	60,32	12,23	67,87	13,62
38	80,77	14,52	82,36	10,38
39	105,96	11,87	73,05	9,92
40	72,7	11,57	89,51	10,61
41	63,85	9,92	75,05	10,83
42	96,32	10,97	86,66	10,36
43	99,41	8,46	70,07	10,67
44	75,58	13,64	62,34	11,81
45	104,87	10,47	64,78	9,38
46	96,06	13,22	71,72	10,06
47	96,83	11,99	61,97	11,25
48	59,64	10,19	77,75	8,48
49	102,03	10,05	84,2	10,69
50	80,62	11,33	59,01	9,57

Anexo 04: Cantidad de Insighs

Tabla 22:

Cantidad de Insighs

N°	CI Facebook	CI Facebook	CI Instagram	CI Instagram
Ejecución	Antes	Después	Antes	Después
01	22	59	16	48
02	19	53	13	42
03	23	61	12	40
04	27	68	17	49
05	18	51	19	53
06	18	51	18	51
07	27	68	11	38
08	23	61	13	42
09	17	49	16	48
10	22	59	18	51
11	17	49	13	42
12	17	49	14	44
13	21	57	10	37
14	10	37	10	37
15	11	38	18	51
16	17	49	20	55
17	14	44	14	44
18	21	57	19	53
19	15	46	16	48
20	12	40	12	40
21	27	68	16	48
22	18	51	21	57
23	20	55	14	44
24	12	40	21	57
25	17	49	4	25
26	20	55	18	51
27	14	44	15	46

28	21	57	13	42
29	16	48	15	46
30	18	51	7	31
31	16	48	14	44
32	29	72	16	48
33	19	53	20	55
34	14	44	12	40
35	24	62	11	38
36	13	42	12	40
37	21	57	18	51
38	10	37	16	48
39	13	42	12	40
40	20	55	17	49
41	23	61	15	46
42	20	55	18	51
43	19	53	12	40
44	18	51	13	42
45	12	40	13	42
46	16	48	9	35
47	17	49	16	48
48	25	64	16	48
49	21	57	15	46
50	11	38	14	44

Anexo 05: Precisión en el Análisis

Tabla 23:

Precisión en el Análisis

N°	PA Facebook	PA Facebook	PA Instagram	PA Instagram
Ejecución	Antes	Después	Antes	Después
01	62	100	59	100
02	59	100	56	100
03	63	100	55	100
04	67	100	60	100
05	58	100	62	100
06	58	100	61	100
07	67	100	54	99
08	63	100	56	100
09	57	100	59	100
10	62	100	61	100
11	57	100	56	100
12	57	100	57	100
13	61	100	53	98
14	50	92	53	98
15	51	94	61	100
16	57	100	63	100
17	54	99	57	100
18	61	100	62	100
19	55	100	59	100
20	52	96	55	100
21	67	100	59	100
22	58	100	64	100
23	60	100	57	100
24	52	96	64	100
25	57	100	47	86
26	60	100	61	100
27	54	99	58	100

28	61	100	56	100
29	56	100	58	100
30	58	100	50	92
31	56	100	57	100
32	69	100	59	100
33	59	100	63	100
34	54	99	55	100
35	64	100	54	99
36	53	98	55	100
37	61	100	61	100
38	50	92	59	100
39	53	98	55	100
40	60	100	60	100
41	63	100	58	100
42	60	100	61	100
43	59	100	55	100
44	58	100	56	100
45	52	96	56	100
46	56	100	52	96
47	57	100	59	100
48	65	100	59	100
49	61	100	58	100
50	51	94	57	100

Anexo 06: Comparación de Modelos de Análisis de Sentimientos

Tabla 24:

Comparación de Modelos de Análisis de Sentimientos

N° Ejecución	Modelo	Resultado	Modelo	Resultado	Modelo	Resultado
01	Naïve				BERT +	
	Bayes	TN	SVM	TN	LSTM	FN
02	Naïve				BERT +	
	Bayes	TP	SVM	TP	LSTM	TN
03	Naïve				BERT +	
	Bayes	FN	SVM	TN	LSTM	TP
04	Naïve				BERT +	
	Bayes	FP	SVM	FP	LSTM	TP
05	Naïve				BERT +	
	Bayes	FP	SVM	TP	LSTM	TN
06	Naïve				BERT +	
	Bayes	FN	SVM	TP	LSTM	TP
07	Naïve				BERT +	
	Bayes	FN	SVM	TP	LSTM	TN
08	Naïve				BERT +	
	Bayes	TP	SVM	TN	LSTM	TN
09	Naïve				BERT +	
	Bayes	TN	SVM	TP	LSTM	TN
10	Naïve				BERT +	
	Bayes	TN	SVM	FN	LSTM	TP
11	Naïve				BERT +	
	Bayes	TN	SVM	TP	LSTM	TN
12	Naïve				BERT +	
	Bayes	TN	SVM	TP	LSTM	FN
13	Naïve				BERT +	
	Bayes	TN	SVM	TN	LSTM	TN
14	Naïve	TP	SVM	TN	BERT +	TN

	Bayes				LSTM	
	Naïve				BERT +	
15	Bayes	TP	SVM	TN	LSTM	TN
	Naïve				BERT +	
16	Bayes	TN	SVM	FN	LSTM	TP
	Naïve				BERT +	
17	Bayes	FN	SVM	TP	LSTM	TP
	Naïve				BERT +	
18	Bayes	TP	SVM	TP	LSTM	TP
	Naïve				BERT +	
19	Bayes	FN	SVM	FN	LSTM	TN
	Naïve				BERT +	
20	Bayes	FN	SVM	TP	LSTM	TN
	Naïve				BERT +	
21	Bayes	TP	SVM	FP	LSTM	TN
	Naïve				BERT +	
22	Bayes	TN	SVM	TP	LSTM	TP
	Naïve				BERT +	
23	Bayes	TN	SVM	FP	LSTM	TN
	Naïve				BERT +	
24	Bayes	TP	SVM	FP	LSTM	FN
	Naïve				BERT +	
25	Bayes	TN	SVM	TN	LSTM	TP
	Naïve				BERT +	
26	Bayes	TP	SVM	TN	LSTM	TN
	Naïve				BERT +	
27	Bayes	TN	SVM	TN	LSTM	TP
	Naïve				BERT +	
28	Bayes	TN	SVM	TP	LSTM	TN
	Naïve				BERT +	
29	Bayes	TP	SVM	TN	LSTM	TP
	Naïve				BERT +	
30	Bayes	TP	SVM	TN	LSTM	TN

31	Naïve				BERT +	
	Bayes	TP	SVM	TP	LSTM	TN
32	Naïve				BERT +	
	Bayes	TN	SVM	TP	LSTM	FP
33	Naïve				BERT +	
	Bayes	TN	SVM	TN	LSTM	TP
34	Naïve				BERT +	
	Bayes	TN	SVM	TP	LSTM	TN
35	Naïve				BERT +	
	Bayes	TN	SVM	FN	LSTM	TN
36	Naïve				BERT +	
	Bayes	FP	SVM	FN	LSTM	TP
37	Naïve				BERT +	
	Bayes	FP	SVM	TN	LSTM	TP
38	Naïve				BERT +	
	Bayes	TN	SVM	TN	LSTM	TP
39	Naïve				BERT +	
	Bayes	TN	SVM	FN	LSTM	TP
40	Naïve				BERT +	
	Bayes	FN	SVM	TP	LSTM	TN
41	Naïve				BERT +	
	Bayes	FP	SVM	TN	LSTM	TP
42	Naïve				BERT +	
	Bayes	TP	SVM	TN	LSTM	TN
43	Naïve				BERT +	
	Bayes	FN	SVM	TN	LSTM	TP
44	Naïve				BERT +	
	Bayes	TN	SVM	TN	LSTM	TN
45	Naïve				BERT +	
	Bayes	TP	SVM	TN	LSTM	TN
46	Naïve				BERT +	
	Bayes	TP	SVM	TP	LSTM	TP
47	Naïve	TP	SVM	TN	BERT +	TN

48	Bayes	TN	SVM	TP	LSTM	TP
	Naïve				BERT +	
49	Bayes	TP	SVM	TP	LSTM	TP
	Naïve				BERT +	
50	Bayes	TP	SVM	TN	LSTM	TP
	Naïve				BERT +	

Anexo 07: Diagrama de la Metodología CRISP – SNA

Figura 62:

Diagrama de Metodología CRISP - SNA

METODOLOGÍA CRISP - SNA

