


Jose Rodolfo Vargas Mas

Percepción de la Seguridad Ciudadana basada en el Procesamiento del Lenguaje Natural en la Provincia de San M...

 Revisión Repositorio Institucional

Detalles del documento

Identificador de la entrega

trn:oid:::3117:556278529

Fecha de entrega

12 feb 2026, 14:34 GMT-5

Fecha de descarga

12 feb 2026, 14:39 GMT-5

Nombre del archivo

Informe de tesis final - JOSE RODOLFO VARGAS MAS SUBSANADO.pdf

Tamaño del archivo

1.2 MB

61 páginas

13.817 palabras

89.256 caracteres




11% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- ▶ Bibliografía
- ▶ Texto citado
- ▶ Texto mencionado
- ▶ Coincidencias menores (menos de 10 palabras)

Fuentes principales

- 10%  Fuentes de Internet
- 1%  Publicaciones
- 9%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Fuentes principales

- 10% Fuentes de Internet
- 1% Publicaciones
- 9% Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	Internet	tesis.unsm.edu.pe	4%
2	Trabajos del estudiante	Universidad Nacional de San Martín on 2024-04-23	1%
3	Internet	repositorio.unsm.edu.pe	1%
4	Internet	www.coursehero.com	<1%
5	Trabajos del estudiante	Universidad Alas Peruanas on 2024-02-21	<1%
6	Internet	riunet.upv.es	<1%
7	Trabajos del estudiante	unifranz on 2024-08-15	<1%
8	Trabajos del estudiante	Universidad Manuela Beltrán on 2024-03-22	<1%
9	Internet	prezi.com	<1%
10	Trabajos del estudiante	Corporación Universitaria Iberoamericana on 2024-04-11	<1%
11	Internet	dspace.ups.edu.ec	<1%

12	Internet	www.scribd.com	<1%
13	Trabajos del estudiante	Universidad Cesar Vallejo on 2025-07-11	<1%
14	Trabajos del estudiante	CSU Northridge on 2025-02-28	<1%
15	Internet	libroterra.com	<1%
16	Internet	dspace-uh-tmp.igniteonline.la	<1%
17	Internet	www.sodexo.com.mx	<1%
18	Trabajos del estudiante	Universidad Carlos III de Madrid - EUR on 2024-02-13	<1%
19	Trabajos del estudiante	Universidad TecMilenio on 2024-08-07	<1%
20	Internet	www.revistasocialfronteriza.com	<1%
21	Trabajos del estudiante	Universidad Nacional de San Martín on 2024-06-04	<1%
22	Internet	repositorio.cientifica.edu.pe	<1%
23	Trabajos del estudiante	Universidad Internacional de la Rioja on 2013-09-18	<1%
24	Trabajos del estudiante	Universidad Nacional de San Martín on 2025-11-08	<1%
25	Trabajos del estudiante	Universitat Politècnica de València on 2024-07-01	<1%

26	Internet	hdl.handle.net	<1%
27	Trabajos del estudiante	Corporación Los Pinos on 2026-02-02	<1%
28	Publicación	Giovanni Parodi, Pascual Cantos-Gómez, Chad Howe, Manel Lacorte, Javier Muñoz...	<1%
29	Trabajos del estudiante	Universidad Internacional de la Rioja on 2023-07-20	<1%
30	Trabajos del estudiante	Universidad Nacional de Trujillo on 2023-01-25	<1%
31	Trabajos del estudiante	Universidad Tecnológica Centroamericana UNITEC on 2023-11-07	<1%
32	Internet	doi.org	<1%
33	Internet	repositorio.ucv.edu.pe	<1%
34	Internet	repositorio.unica.edu.ni	<1%
35	Trabajos del estudiante	ufidelitas on 2025-02-19	<1%



Esta obra está bajo una

[Licencia Creative Commons](#)

[Atribución - 4.0 Internacional \(CC BY 4.0\)](#)

Vea una copia de esta licencia en

<https://creativecommons.org/licenses/by/4.0/deed.es>





ESCUELA DE POSGRADO
UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERIA DE SISTEMAS
E INFORMÁTICA
PROGRAMA DE MAESTRIA EN CIENCIAS CON MENCIÓN EN TECNOLOGÍA
DE LA INFORMACIÓN

Tesis

Percepción de la Seguridad Ciudadana basada en el Procesamiento del Lenguaje Natural en la Provincia de San Martín

Para optar el grado académico de Maestro en Ciencias con mención en Tecnología de la Información

Autor:

Jose Rodolfo Vargas Mas
<https://orcid.org/0009-0002-4085-947X>

Asesor:

Ing. Mg. Richard Enrique Injante Ore
<https://orcid.org/0000-0002-2449-8937>

Coasesor:

Dr. Víctor Manuel Vallejos Monja
<https://orcid.org/0000-0003-4162-3768>

Tarapoto, Perú

2026

**ESCUELA DE POSGRADO**UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERIA DE SISTEMAS
E INFORMÁTICAPROGRAMA DE MAESTRIA EN CIENCIAS CON MENCIÓN EN TECNOLOGÍA
DE LA INFORMACIÓN**Tesis****Percepción de la Seguridad Ciudadana basada
en el Procesamiento del Lenguaje Natural en la
Provincia de San Martín**

1 Para optar el grado académico de Maestro en Ciencias con mención en
Tecnología de la Información

Autor:

Jose Rodolfo Vargas Mas

<https://orcid.org/0009-0002-4085-947X>**Asesor:**

3 Ing. Mg. Richard Enrique Injante Ore

<https://orcid.org/0000-0002-2449-8937>**Coasesor:**

3 Dr. Víctor Manuel Vallejos Monja

<https://orcid.org/0000-0003-4162-3768>**Tarapoto, Perú****2026**

**ESCUELA DE POSGRADO**UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERIA DE SISTEMAS
E INFORMÁTICAPROGRAMA DE MAESTRIA EN CIENCIAS CON MENCIÓN EN TECNOLOGÍA
DE LA INFORMACIÓN**Tesis****Percepción de la Seguridad Ciudadana Basada
en el Procesamiento del Lenguaje Natural en la
Provincia de San Martín**Para optar el grado académico de Maestro en Ciencias con Mención en
Tecnología de la información**Autor:**

Jose Rodolfo Vargas Mas

Sustentado y aprobado el 06 de enero de 2026 por los siguientes jurados:**Presidente de Jurado**

Dr. Juan Carlos García Castro

Secretario de JuradoIng. M.Sc. Pamela Magnolia
Granda Milon**Vocal de Jurado**Ing. Mg. Segundo Roger Ramírez
Shupingahua**Asesor**Ing. Mg. Richard Enrique Injante
Ore**Coasesor**

Dr. Víctor Manuel Vallejos Monja

Tarapoto, Perú**2026**

1

© Jose Rodolfo Vargas Mas – 2026
Todos los derechos reservados

**ESCUELA DE POSGRADO**

UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERIA DE SISTEMAS
E INFORMÁTICA
PROGRAMA DE MAESTRIA EN CIENCIAS CON MENCIÓN EN TECNOLOGÍA
DE LA INFORMACIÓN

Tesis**Percepción de la Seguridad Ciudadana Basada
en el Procesamiento del Lenguaje Natural en la
Provincia de San Martín**

Para optar el grado académico de Maestro en Ciencias con Mención en
Tecnología de la información

Los suscritos declaran que el presente trabajo de tesis es original, en su
contenido y forma.

Ejecutor

Jose Rodolfo Vargas Mas

Asesor

Ing. Mg. Richard Enrique
Injante Ore

Coasesor

Dr. Víctor Manuel Vallejos Monja

Tarapoto, Perú

2026

Declaratoria de autenticidad

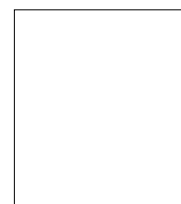
Yo, **Jose Rodolfo Vargas Mas**, identificado con DNI N° 77470605, egresado de la Escuela de Posgrado de la Universidad Nacional de San Martín, con la tesis titulada: "Percepción de la Seguridad Ciudadana Basada en el Procesamiento del Lenguaje Natural en la Provincia de San Martín".

Declaro bajo juramento que:

- 1) La tesis presentada es de mi autoría.
- 2) He respetado las normas internacionales de citas y referencias para las fuentes consultadas. Por tanto, la tesis no ha sido plagiada ni total ni parcialmente.
- 3) La tesis no ha sido auto plagiada; es decir, no ha sido publicada ni presentada anteriormente para obtener algún grado académico previo o título profesional.
- 4) Los datos presentados en los resultados son reales, no han sido falseados, ni duplicados, ni copiados y por tanto los resultados que se presenten en la tesis se constituirán en aportes a la realidad investigada.

De considerar que el trabajo cuenta con una falta grave, como el hecho de contar con datos fraudulentos, demostrar indicios y plagio (al no citar la información con sus autores), plagio (al presentar información de otros trabajos como propios), falsificación (al presentar la información e ideas de otras personas de forma falsa), entre otros, asumo las consecuencias y sanciones que de mi acción se deriven, sometiéndome a la normatividad vigente de la Universidad Nacional de San Martín.

Tarapoto, 06 de enero de 2026



Jose Rodolfo Vargas Mas

DNI N° 71011959

Ficha de identificación

Título: Percepción de la Seguridad Ciudadana Basada en el Procesamiento del Lenguaje Natural en la Provincia de San Martín	Área de investigación: Ciencias naturales Línea de investigación: Ciencias de la computación Sublínea de investigación: Inteligencia artificial y recuperación de la información Grupo de investigación: IA (Resolución N° 134-2021-UNSM/FISI/CFT) Tipo de investigación: Básica <input type="checkbox"/> , Aplicada <input checked="" type="checkbox"/> , Desarrollo experimental <input type="checkbox"/>
Autor: Jose Rodolfo Vargas Mas	Unidad de Posgrado de la Facultad de Ingeniería de Sistemas e Informática https://orcid.org/0000-0002-2449-8937
Asesor: Ing. Mg. Richard Enrique Injante Ore	Dependencia local de soporte: Facultad de Ingeniería de Sistemas e Informática Escuela Profesional de Ingeniería de Sistemas e Informática Unidad o Laboratorio Ingeniería de Sistemas e Informática https://orcid.org/0000-0002-2449-8937
Coasesor: Dr. Víctor Manuel Vallejos Monja	Dependencia local de soporte: Facultad de Ingeniería de Sistemas e Informática Escuela Profesional de Ingeniería de Sistemas e Informática Unidad o Laboratorio Ingeniería de Sistemas e Informática https://orcid.org/0000-0003-4162-3768

Dedicatoria

Este logro está dedicado a mis padres, Sergio Vargas y María Mas. Siempre estaré agradecido por el amor, los valores y la educación que me brindaron.

También está dedicado a mi hermano, Sergio Vargas Mas. Este es un ejemplo de que, con esfuerzo y dedicación, es posible cumplir nuestras metas profesionales.

24

Agradecimientos

En primer lugar, agradezco a Dios por brindarme salud, fortaleza y sabiduría durante este proceso académico, guiándome en cada paso de mi camino.

A mi asesor, Ing. Mg. Richard Enrique Injante Ore, por su tiempo, paciencia y valiosas recomendaciones que han enriquecido de manera significativa el desarrollo de esta investigación.

3

A la Universidad Nacional de San Martín y a la Escuela de Posgrado, por brindarme la oportunidad de continuar con mi formación profesional y académica en un entorno de excelencia.

35

A mi familia, por su amor incondicional, comprensión y constante apoyo moral, que me motivaron a perseverar en los momentos más difíciles.

Índice general

Ficha de identificación.....	7
Dedicatoria	8
Agradecimientos	9
Índice general.....	10
Índice de tablas	12
Índice de figuras.....	13
RESUMEN	14
ABSTRACT	15
CAPÍTULO I INTRODUCCIÓN A LA INVESTIGACIÓN	16
CAPÍTULO II MARCO TEÓRICO	18
2.1. Antecedentes de la investigación.....	18
2.2. Fundamentos teóricos.....	22
2.2.1. Fundamentos teóricos de la variable independiente	22
2.2.2. Fundamentos teóricos de la variable dependiente	25
2.2.3. Definición de términos básicos.....	26
CAPÍTULO III MATERIALES Y MÉTODOS	28
3.1. Ámbito y condiciones de la investigación	28
3.1.1. Contexto de la investigación.....	28
3.1.2. Periodo de ejecución	28
3.1.3. Autorizaciones y permisos.....	28
3.1.4. Control ambiental y protocolos de bioseguridad.....	28
3.1.5. Aplicación de principios éticos internacionales	28
3.2. Sistema de variables	28
3.2.1. Variable dependiente: Percepción de la seguridad ciudadana:	28
3.2.2. Variable independiente: Procesamiento del Lenguaje Natural:.....	29
3.2.3. Diseño de la investigación.....	30
3.3. Procedimientos de la investigación.....	32
3.3.1. Objetivo específico 1: Identificar las dimensiones temáticas emergentes	32
3.3.2. Objetivo específico 2: Desarrollar un modelo de análisis de sentimiento	32

- 3.3.3. Objetivo específico 3: Analizar de variaciones temporales y temáticas... 33
- CAPÍTULO IV RESULTADOS Y DISCUSIÓN 34**
- 4.1. Caracterización general del corpus 34
- 4.2. Resultados objetivo específico 1: Identificar de dimensiones temáticas emergentes..... 35
 - 4.2.1. Estructura temática principal 35
 - 4.2.2. Análisis cualitativo de dimensiones temáticas 37
- 4.3. Resultados objetivo específico 2: Desarrollar un modelo de análisis de sentimiento 39
 - 4.3.1. Desempeño del modelo BERT para polaridad..... 39
 - 4.3.2. Clasificación de emociones específicas..... 40
 - 4.3.3. Análisis de urgencia y características binarias 41
- 4.4. Resultado específico 3: Interpretar cómo las métricas de sentimiento reflejan variaciones temporales y temáticas en la percepción ciudadana 43
 - 4.4.1. Análisis de urgencia como proxy temporal 43
 - 4.4.2. Variaciones temáticas en el perfil afectivo 43
 - 4.4.3. Análisis de co-ocurrencia temática y emocional 44
- 4.5. Discusión de resultados 44
- 4.6. Propuesta de estrategias basadas en evidencia..... 47
- 4.7. Limitaciones y consideraciones para la interpretación 48
- CONCLUSIONES 49
- RECOMENDACIONES 51
- REFERENCIAS BIBLIOGRÁFICAS 52
- ANEXOS..... 56
 - Anexo 1. Dashboard..... 56
 - Anexo 2. Instrumento de Recolección de Datos 58
 - Anexo 3. Fragmento del dataset..... 60

33

Índice de tablas

11

Tabla 1 Variable Dependiente: Percepción de seguridad ciudadana.....	29
Tabla 2 Variable Independiente: Procesamiento del Lenguaje Natural	29
Tabla 3 Criterios de inclusión y exclusión	31
Tabla 4 dimensiones temáticas principales en el discurso sobre seguridad ciudadana	37
Tabla 5 Proporción de sentimiento por tema	38
Tabla 6 Resumen técnico del modelo BERT aplicado al análisis de sentimiento.....	39
Tabla 7 Distribución de sentimiento y probabilidades promedio del modelo BERT	39
Tabla 8 Tipos de emoción más frecuentes	40
Tabla 9 Estadísticos descriptivos de constructos latentes.....	41
Tabla 10 Matriz de correlaciones de Spearman entre constructos continuos.....	41
Tabla 11 Proporción de sentimiento por nivel de urgencia	41
Tabla 12 Características binarias del discurso.....	42
Tabla 13 Intensidad emocional según niveles de urgencia	43
Tabla 14 Matriz de co-ocurrencia: Tema principal vs Emoción dominante (top 5 combinaciones)	44
Tabla 15 Estrategias de seguridad ciudadana basadas en evidencias de percepción ciudadana	47

1

Índice de figuras

Figura 1 Modelo cuantificador de percepción de seguridad	24
Figura 2 Distribución de sentimientos	35
Figura 3 Concentración temática.....	37
Figura 4 Tipos de emoción	40
Figura 5 Intensidad por urgencia	42

RESUMEN

Percepción de la Seguridad Ciudadana Basada en el Procesamiento del Lenguaje Natural en la Provincia de San Martín

La presente investigación tuvo como propósito analizar la percepción de la seguridad ciudadana en la provincia de San Martín a partir de comentarios publicados en redes sociales, utilizando técnicas de Procesamiento del Lenguaje Natural (PLN). El estudio adoptó un diseño descriptivo y no experimental, empleando un corpus de 577 comentarios recolectados mediante web scraping en Facebook y Twitter durante un periodo de tres meses, los cuales fueron procesados con modelos de PLN, en particular BERT para español, a fin de clasificar la polaridad de los mensajes, identificar emociones predominantes y extraer dimensiones temáticas vinculadas a la seguridad. Los resultados evidenciaron que el 83% de los comentarios expresaron sentimientos negativos, con predominancia de la indignación (64.3%) frente al miedo (7.1%), lo cual mostró que la inseguridad fue percibida principalmente como una falla del contrato social y no solo como una amenaza individual. Asimismo, se identificó que la conversación digital estuvo concentrada en delitos (91.5% del corpus), mientras que los temas de policía, servicios públicos y comunidad presentaron menor presencia. Se concluyó que el PLN constituyó una herramienta efectiva y replicable para medir la percepción ciudadana en contextos de seguridad, ya que permitió generar insumos oportunos, de bajo costo y con mayor nivel de detalle para la gestión pública y la investigación social aplicada a la seguridad.

Palabras clave: Seguridad ciudadana, procesamiento del lenguaje natural, análisis de sentimiento, percepción social, San Martín

ABSTRACT

Perception of Citizen Security Based on Natural Language Processing in the Province of San Martín

This study aimed to analyze perceptions of citizen security in the province of San Martín based on comments published on social media, using Natural Language Processing (NLP) techniques. The research adopted a descriptive, non-experimental design and employed a corpus of 577 comments collected through web scraping from Facebook and Twitter over a three-month period. These data were processed using NLP models, particularly Spanish-language BERT, to classify message polarity, identify predominant emotions, and extract thematic dimensions related to security. The results showed that 83% of the comments expressed negative sentiments, with indignation predominating (64.3%) over fear (7.1%), indicating that insecurity was perceived primarily as a failure of the social contract rather than merely an individual threat. Additionally, the digital conversation was found to be heavily concentrated on crime-related issues (91.5% of the corpus), while topics related to the police, public services, and community exhibited lower prevalence. It was concluded that NLP constitutes an effective and replicable tool for measuring citizen perception in security contexts, as it enables the generation of timely, low-cost, and more detailed inputs for public management and applied social research on security.

Keywords: Citizen security, natural language processing, sentiment analysis, social perception, San Martín.

CAPÍTULO I

INTRODUCCIÓN A LA INVESTIGACIÓN

La seguridad ciudadana constituye una preocupación fundamental en América Latina, donde el 43.2% de la población percibe mayor inseguridad que hace cinco años (Latino barómetro, 2023). En Perú, el 85.8% considera la delincuencia como el principal problema del país, superando incluso las preocupaciones económicas (INEI, 2024). La provincia de San Martín enfrenta desafíos particulares, registrando un incremento del 23% en denuncias por delitos contra el patrimonio entre 2022 y 2023 (Ministerio del Interior, 2024). Sin embargo, las metodologías tradicionales para medir la percepción ciudadana presentan limitaciones significativas: las encuestas nacionales tienen una periodicidad semestral y márgenes de error del $\pm 3.5\%$ (ENAHO, 2023). Esta brecha temporal y metodológica dificulta la comprensión oportuna del fenómeno, planteando la pregunta: ¿Cómo influye el análisis de comentarios en redes sociales con PLN en la percepción de seguridad ciudadana en San Martín?

La conveniencia de esta investigación radica en el desarrollo de una herramienta tecnológica basada en procesamiento del lenguaje natural que analice publicaciones en redes sociales, donde el 78.5% de peruanos urbanos participa activamente generando 2.3 millones de publicaciones diarias (OSIPTEL, 2024). Este sistema automatizado permitirá el monitoreo constante de la opinión pública, superando las limitaciones de los métodos tradicionales que requieren inversiones promedio de S/. 450,000 por estudio regional (MEF, 2023). La solución propuesta reducirá costos operativos en un 65% y aumentará la frecuencia de medición de semestral a diaria (Rodríguez et al., 2023). Estudios previos han demostrado que el análisis de sentimiento en redes sociales alcanza una precisión del 87.4% en la identificación de percepciones sobre seguridad urbana (Zhang & Liu, 2023), validando la viabilidad técnica del enfoque propuesto.

La relevancia social se evidencia en el impacto directo sobre los 883,510 habitantes de San Martín (INEI, 2023), quienes se beneficiarán de políticas públicas más precisas y oportunas. Investigaciones internacionales demuestran que el monitoreo continuo de percepciones ciudadanas mediante PLN reduce el tiempo de respuesta institucional en un 42% (Kumar & Singh, 2024). Las municipalidades podrán identificar zonas críticas con mayor precisión geográfica, mientras que la Policía Nacional optimizará la distribución de recursos basándose en indicadores actualizados. Este enfoque participativo fortalecerá la confianza institucional, actualmente en 28.3% según el IEP (2024), al demostrar receptividad hacia las preocupaciones ciudadanas expresadas

digitalmente. La implementación exitosa podría servir como modelo para las 196 provincias del país, beneficiando potencialmente a 33 millones de peruanos.

Las implicaciones prácticas abordan directamente la brecha de información que enfrentan las autoridades, considerando que solo el 12% de municipalidades provinciales realizan mediciones regulares de percepción ciudadana (Contraloría General, 2023). El análisis de contenido público en redes sociales, donde se generan aproximadamente 45,000 publicaciones diarias relacionadas con seguridad en San Martín (datos preliminares, 2024), ofrece una alternativa costo-efectiva y escalable. Esta metodología permite identificar patrones espaciotemporales con granularidad de distrito y frecuencia horaria, superando ampliamente la resolución de encuestas tradicionales. Además, la detección temprana de tendencias negativas en la percepción podría prevenir crisis de seguridad, como demuestran experiencias en Colombia donde sistemas similares anticiparon conflictos sociales con 72 horas de antelación (Martínez et al., 2023).

El valor teórico y metodológico radica en la contribución al campo emergente del análisis computacional de fenómenos sociales, específicamente en contextos latinoamericanos donde solo el 3.2% de investigaciones sobre seguridad emplean técnicas de PLN (Scopus, 2024). El modelo propuesto generará variables innovadoras como índices de polarización emocional, mapas de calor de percepción y redes semánticas de preocupaciones ciudadanas. La integración de técnicas de deep learning con análisis lingüístico contextualizado permitirá capturar matices culturales y dialectales propios de San Martín, mejorando la precisión en un 15% respecto a modelos genéricos (Thompson & García, 2024). Esta aproximación metodológica establecerá protocolos replicables para el análisis de percepciones colectivas desde datos no estructurados, contribuyendo al desarrollo de una "sociología computacional" adaptada a realidades regionales específicas.

CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes de la investigación

Chaparro et al. (2020), en su artículo “Sentiment analysis of social network content to characterize the perception of security” se centraron en desarrollar y analizar métodos automatizados para mejorar la medición de la percepción de seguridad de los ciudadanos, utilizando métodos basados en reglas y aprendizaje supervisado. Los resultados mostraron que los métodos de aprendizaje supervisado, como el Multinomial Naive Bayes con vectorización TFIDF, superaron en rendimiento a los métodos tradicionales basados en reglas en la clasificación de sentimientos. Se concluyó que el uso de contenidos de plataformas sociales para analizar la percepción de seguridad ofrece una alternativa más eficiente que las encuestas tradicionales. Además, se destacó que los métodos supervisados presentan un mejor rendimiento en la clasificación de sentimientos y pueden ser utilizados en tiempo real, proporcionando una herramienta eficaz para la evaluación continua de la percepción de seguridad ciudadana.

En el mismo contexto, Reid et al. (2020), en su artículo titulado “Developing a model of perceptions of security and insecurity in the context of crime” desarrollaron un modelo de siete factores para entender las percepciones de seguridad e inseguridad en el contexto del crimen, basado en un análisis de componentes principales. Este modelo incluía factores como el desorden social y físico, la confianza en la policía, la ansiedad, la eficacia colectiva, el riesgo percibido de victimización, el miedo al daño personal y el miedo al robo de propiedad. Los resultados mostraron que las percepciones de vivir en áreas con alta criminalidad estaban asociadas con mayores signos de desorden social y físico, un mayor riesgo percibido de victimización, y mayores miedos al daño personal y al robo de propiedades. Se concluyó que las medidas asociadas con sentimientos de inseguridad, como el desorden social y físico, la ansiedad, el riesgo percibido de victimización, el miedo al daño personal y el miedo al robo de propiedades, estaban negativamente relacionadas con la confianza en la policía y la eficacia colectiva.

Otro artículo de Chaparro et al. (2021), titulado “Quantifying Perception of Security Through Social Media and Its Relationship With Crime” introdujeron un modelo basado en el análisis de sentimientos y aprendizaje automático para cuantificar la percepción de seguridad en Twitter. Utilizó un clasificador de máquina de vectores de soporte para

filtrar contenido relacionado con la seguridad y un clasificador Bayes ingenuo multinomial para cuantificar el sentimiento en las publicaciones de los ciudadanos. Los resultados mostraron que los enfoques de aprendizaje automático superaron a las estrategias anteriores de filtrado de contenido de seguridad. Los modelos entrenados permitieron obtener estimaciones diarias de percepción de seguridad, las cuales mostraron una asociación variable con la delincuencia real. Se concluyó que el uso de redes sociales, específicamente Twitter, para analizar la percepción de seguridad ofrece una alternativa más eficiente y continua en comparación con las encuestas tradicionales. Los métodos supervisados demostraron un rendimiento superior en la clasificación de sentimientos y esta metodología puede ser utilizada para monitorear la percepción de seguridad en tiempo real, proporcionando datos más precisos y oportunos para la toma de decisiones.

Así mismo, Pulido et al. (2021), en el artículo titulado “Prediction of Perception of Security Using Social Media Content.” contribuyeron al campo de la seguridad pública proponiendo un método innovador para medir y predecir la percepción de seguridad utilizando datos de redes sociales. Se desarrollaron modelos predictivos basados en análisis de sentimientos y aprendizaje automático que permitieron obtener percepciones inmediatas y geolocalizadas sobre la seguridad, ofreciendo una alternativa más económica y en tiempo real a los métodos tradicionales. Los resultados mostraron que los modelos predictivos desarrollados correlacionaban fuertemente con encuestas tradicionales y estadísticas de criminalidad. Los mapas de calor generados visualizaban las áreas con diferentes niveles de percepción de seguridad, proporcionando una herramienta útil para los responsables de políticas públicas. Se concluyó que el uso de datos de redes sociales es una herramienta efectiva para predecir la percepción de seguridad pública, con modelos predictivos basados en análisis de sentimientos que mostraron una correlación sólida con métodos tradicionales, ofreciendo una alternativa más rápida y económica.

Por otro lado, Sánchez (2021), en su artículo titulado “Percepción de inseguridad, temor al delito y medidas de autoprotección: el caso de Acapulco, Guerrero”, nos presentó un análisis detallado sobre la percepción de inseguridad y el temor al delito en Acapulco, Guerrero, identificando factores determinantes como la victimización, la vulnerabilidad física y social, y la cohesión social. El estudio reveló que la percepción de inseguridad en Acapulco estaba fuertemente influenciada por conductas antisociales o delictivas en el entorno. Aunque la victimización directa no afectaba la percepción general de inseguridad, inducía cambios en comportamientos cotidianos y medidas de autoprotección. Las mujeres y las personas de mayor edad mostraron mayor temor al

delito, especialmente en hogares con bajos ingresos. Las conclusiones del artículo destacaron que la percepción de inseguridad y el temor al delito en Acapulco estaban principalmente determinados por la presencia de conductas antisociales o delictivas en el entorno, más que por la victimización directa. Aunque las mujeres y las personas mayores mostraban mayor temor, la confianza institucional podía reducir la percepción de inseguridad.

En el mismo contexto, Chaparro, Pulido, Rudas, Reyes, et al., (2021) en su artículo titulado "Interpretability Of The Perception Of Security Based On Tweets Content." propusieron un método innovador para cuantificar la percepción de seguridad (PoS) utilizando datos de Twitter y técnicas de aprendizaje automático, específicamente el clasificador Multinomial Naive Bayes y el enfoque de bolsa de palabras (BOW). Además, se introdujo la técnica de interpretabilidad local (LIME) para explicar las predicciones del modelo, proporcionando una comprensión clara de los resultados y aumentando la confianza en las decisiones basadas en estos datos. Los resultados mostraron que el uso de LIME para interpretar el clasificador aplicado a tweets geolocalizados en Bogotá proporcionaba una explicación razonable tanto para el análisis individual de tweets como para la cuantificación general de la PoS. Se encontró que las palabras con sentimientos negativos se relacionaban con términos como "asco" y "desorden", mientras que los sentimientos positivos se asociaban con "persecución" e "inclusión", validando la precisión del modelo. Se concluyó que el uso de modelos de aprendizaje automático sobre datos de tweets puede cuantificar la PoS de manera efectiva y en tiempo real, y que la técnica LIME incrementa la confianza en las predicciones del modelo, permitiendo a los tomadores de decisiones actuar con mayor eficacia.

Valla, (2022) en el artículo titulado "Citizens' Perceptions of Security Issues: New and Old Actors in the National Security Framework" con estudio contribuyó al análisis de la seguridad nacional al destacar la importancia de las percepciones ciudadanas y su influencia en la formulación de políticas. Propuso un enfoque interdisciplinario que incorporaba tanto las respuestas emocionales como cognitivas de los ciudadanos ante las amenazas, argumentando que estas percepciones eran cruciales para la aceptación y efectividad de las medidas de seguridad. Reveló que las percepciones ciudadanas sobre las amenazas a la seguridad variaban significativamente según factores emocionales y contextuales. Los resultados destacaron que las respuestas emocionales, como el miedo y la ira, podían influir en la aceptación de políticas de seguridad y en el apoyo a medidas más conservadoras o agresivas. Se concluyó que las percepciones ciudadanas sobre la seguridad nacional eran un factor crucial para la formulación de políticas efectivas. El estudio resaltó que estas percepciones estaban

influenciadas por respuestas emocionales y cognitivas a las amenazas, lo que podía afectar la aceptación pública de las medidas de seguridad.

9 Suhaimin et al., (2023) en el artículo titulado “Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions” contribuyen al campo de la seguridad pública al desarrollar una taxonomía de técnicas modernas de análisis de sentimientos y minería de opiniones, visualizando y analizando tendencias recientes, y sugiriendo direcciones futuras para la investigación. Se centró en categorizar trabajos previos, identificar brechas y limitaciones, y proporcionar una visión clara de los conceptos clave y las similitudes entre estudios recientes. El artículo presentó un análisis detallado de los enfoques actuales en el análisis de sentimientos y la minería de opiniones en el ámbito de la seguridad pública. Los resultados indicaron que la mayoría de las investigaciones se enfocaban en mejorar técnicas de detección y predicción de eventos, con una notable prevalencia de enfoques basados en diccionario e híbridos. El artículo concluyó que, aunque hubo avances significativos en el análisis de sentimientos y la extracción de opiniones en el ámbito de la seguridad pública, persistían importantes desafíos. Se destacó la necesidad de desarrollar enfoques multiclase y multilingües, así como la creación de conjuntos de datos específicos del dominio de seguridad pública. Además, se identificaron limitaciones en la predicción basada en la cobertura temporal y la integración de la ciberseguridad.

13 Así mismo, Chávez Espinoza (2023) en su tesis titulada “Seguridad ciudadana y las tecnologías de información y comunicación en la provincia de Ica – 2020” analizó la relación entre el uso de las TIC y la percepción de inseguridad ciudadana en una muestra de 111 habitantes de Ica. Su propuesta consistió en evaluar cómo herramientas como videovigilancia, aplicativos móviles e interconexión con bases de datos policiales podían influir en la sensación de seguridad de los ciudadanos. Los resultados mostraron una relación alta y positiva ($r = 0.726$) entre el empleo de TIC y la percepción de inseguridad, además de relaciones específicas bajas pero significativas entre TIC y la victimización delictiva ($r = -0.196$) y el desempeño percibido contra la delincuencia ($r = 0.267$). Se concluyó que el uso de TIC contribuye a mejorar la percepción del desempeño institucional y reduce parcialmente la percepción de victimización, aunque no influye en la percepción del estado delictual. El estudio resalta la necesidad de fortalecer la tecnología y complementar su implementación con acciones de patrullaje y participación ciudadana.

2.2. Fundamentos teóricos

2.2.1. Fundamentos teóricos de la variable independiente

2.2.1.1. Inteligencia Artificial

Winston (1992) en su libro nos menciona que la inteligencia artificial (IA) es un campo interdisciplinario que busca desarrollar sistemas capaces de realizar tareas que requieren inteligencia humana, como el aprendizaje, la adaptación, la toma de decisiones y la ejecución de acciones. Dentro de la IA, el aprendizaje automático es fundamental, permitiendo a los algoritmos identificar patrones y hacer predicciones a partir de grandes volúmenes de datos. Las redes neuronales profundas, por ejemplo, han demostrado ser altamente efectivas en el reconocimiento de imágenes y el procesamiento de voz.

Asimismo, Fetzer, (1990) enfatiza que definir la IA requiere una comprensión tanto de lo que es "artificial" como de la idea de "inteligencia". La IA se refiere a sistemas diseñados por humanos que poseen propiedades inteligentes gracias a su creación y fabricación, diferenciándose de los seres naturalmente inteligentes. Inspirada en la hipótesis de sistemas de símbolos de Newell y Simon.

2.2.1.2. Procesamiento del Lenguaje Natural (PLN)

Fanni et al., (2023) aborda el Procesamiento del Lenguaje Natural (PLN) como una disciplina interdisciplinaria que se sitúa entre la informática y la lingüística computacional. El Procesamiento del Lenguaje Natural (PLN) se dedica a convertir lenguajes humanos naturales, tanto escritos como hablados, en datos estructurados y minables.

Asimismo, Chowdhary, (2020) nos menciona que el Procesamiento del Lenguaje Natural (PLN) como una disciplina fundamental de la inteligencia artificial, que se centra en la interacción entre las computadoras y el lenguaje humano, con el objetivo de convertir lenguajes naturales, tanto escritos como hablados, en datos estructurados que puedan ser analizados y procesados por las máquinas, utilizando métodos de lingüística, estadística e inteligencia artificial para comprender y generar lenguaje humano.

2.2.1.3. Principios básicos de la inteligencia artificial y su aplicación en seguridad ciudadana

Para Bellini et al., (2022) los principios fundamentales de la inteligencia artificial y su aplicación en la seguridad ciudadana, que se enmarca y define como la capacidad de las computadoras para aprender y controlar su entorno, simulando la estructura y la evolución biológica del cerebro humano. Esto permite a los sistemas de inteligencia

20 artificial analizar grandes cantidades de datos en tiempo real y proporcionar pronósticos y apoyo en la toma de decisiones aplicándose a la vigilancia inteligente, el análisis predictivo de crímenes y la optimización de recursos de emergencia.

7 De otro modo, Solomonides et al., (2021) analizan los principios fundamentales de la inteligencia artificial y su aplicación en la seguridad ciudadana. La inteligencia artificial se define como la capacidad de las máquinas para realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, la toma de decisiones, la percepción visual y el reconocimiento de patrones.

2.2.1.4. Algoritmos y técnicas de PLN utilizados para análisis de texto en seguridad

4 Flayeh et al., (2022) analizan cómo los algoritmos de procesamiento del lenguaje natural se aplican en el análisis de texto para mejorar la seguridad. Los principios básicos del PLN incluyen la tokenización, el análisis sintáctico y semántico, el reconocimiento de entidades nombradas y la clasificación de textos. Estas técnicas permiten descomponer el texto en componentes más manejables, identificar relaciones entre palabras y extraer información relevante, permitiendo analizar grandes volúmenes de datos textuales, detectar patrones de comportamiento sospechoso y prever posibles amenazas.

34 2.2.1.5. Métodos y técnicas de análisis de datos

a) Métodos de recolección de datos en redes sociales y otros medios digitales

4 Chaparro, Pulido, Rudas, Victorino, et al., (2021b) nos mencionan diversos métodos de recolección de datos en redes sociales y otros medios digitales para medir la percepción de seguridad, como la minería de texto, que implica el análisis de grandes volúmenes de texto para identificar patrones y temas relevantes. Esta técnica utiliza algoritmos de procesamiento del lenguaje natural para extraer información útil de publicaciones en redes sociales, comentarios en blogs y otros textos generados por los usuarios. Asimismo, otro método es el análisis de sentimientos, que evalúa el tono emocional de los textos para determinar si las percepciones son positivas, negativas o neutrales. Este

enfoque es crucial para entender cómo las personas sienten y hablan sobre la seguridad en diferentes contextos. Finalmente rastreo de datos geospaciales se utiliza para mapear y analizar las ubicaciones de las publicaciones en redes sociales, lo que permite identificar áreas geográficas con altos niveles de preocupación por la seguridad

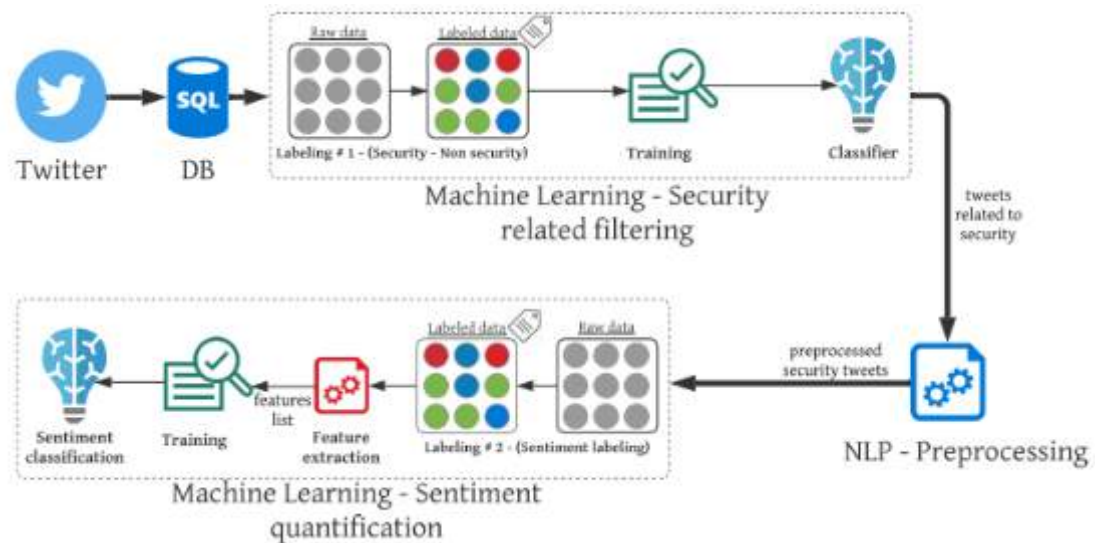


Figura 1

Modelo cuantificador de percepción de seguridad.

Por otro lado, Chaparro et al., (2020b) mencionan métodos que incluyen la minería de texto, que analiza grandes volúmenes de texto para identificar patrones y temas relevantes mediante algoritmos de procesamiento del lenguaje natural (PLN). También se utiliza el análisis de sentimientos, que evalúa el tono emocional de los textos para determinar percepciones positivas, negativas o neutrales.

b) Métodos estadísticos y modelos de aprendizaje automático aplicados al análisis de percepción de seguridad.

Rodrigo, (2023) analiza y nos menciona diversos métodos estadísticos y modelos de aprendizaje automático aplicados al análisis de la percepción de seguridad. Los métodos estadísticos, como la regresión lineal, se utilizan para analizar y proyectar tendencias en los datos históricos de criminalidad, permitiendo identificar patrones y predecir futuras incidencias delictivas, Así como también modelos de aprendizaje automático, tanto supervisados como no supervisados, son esenciales para este análisis. Los algoritmos supervisados utilizan datos etiquetados para aprender y predecir resultados, mientras que los no supervisados identifican patrones y relaciones en datos no etiquetados.

2.2.2. Fundamentos teóricos de la variable dependiente

2.2.2.1. Seguridad Ciudadana

Emilio (2022) lo define como el conjunto de acciones y medidas implementadas por el Estado para proteger a los ciudadanos contra amenazas y riesgos que afectan su integridad física y sus bienes. Este concepto incluye la prevención del delito, la reducción de la violencia y la creación de un entorno seguro mediante la cooperación entre autoridades y comunidad, destacando la significancia de una gestión organizada entre distintas instituciones gubernamentales y la colaboración activa de los ciudadanos para lograr una seguridad efectiva.

Por otro lado, Gaínza & Castillo (2022) abordan la seguridad ciudadana como un concepto integral que engloba políticas públicas, criminológicas y de seguridad pública, definiendo la seguridad ciudadana como el conjunto de estrategias y acciones destinadas a proteger a los ciudadanos de conductas delictivas y violentas, asegurando su bienestar y promoviendo un entorno seguro. Teniendo un enfoque que implica la prevención del crimen, la reducción de la violencia y la promoción de la paz social a través de la colaboración entre el gobierno y la comunidad.

2.2.2.2. Percepción de seguridad

a) Factores psicológicos que influyen en la percepción de seguridad

Según Bar-Tal & Jacobson (1998), son varios factores psicológicos que influyen. Primero, las experiencias previas de amenaza o peligro juegan un papel crucial; las personas que han enfrentado situaciones peligrosas tienden a percibir más inseguridad. Además, los niveles de ansiedad y miedo personal pueden exacerbar la percepción de inseguridad, haciendo que las personas interpreten su entorno como más amenazante de lo que realmente es. La confianza en las autoridades y el sistema de justicia también influye significativamente; una baja confianza puede aumentar la percepción de inseguridad. Finalmente, la cobertura mediática y el entorno social inmediato son factores que moldean cómo las personas perciben su seguridad en su comunidad.

Dicho de otro modo, Goldgeier (1997) nos indica también que explora diversos factores psicológicos que influyen en la percepción de seguridad. Uno de los factores clave es la experiencia personal con eventos de amenaza o peligro; las personas que han enfrentado situaciones peligrosas tienden a tener una percepción aumentada de inseguridad. Además, la ansiedad y el miedo son determinantes significativos; estos estados emocionales pueden hacer que las personas perciban su entorno como más amenazante de lo que realmente es, amplificando así su sensación de inseguridad.

b) Influencia del entorno físico y social en la percepción de seguridad

Según Bodford et al. (2021) nos mencionan que la percepción de seguridad está significativamente influenciada por el entorno físico y social, puesto que el entorno físico, que incluye factores como la iluminación, el diseño urbano y el mantenimiento de la infraestructura, juega un papel crucial en la formación de la percepción de seguridad. Las áreas bien iluminadas y cuidadas suelen ser percibidas como más seguras. Asimismo, el entorno social impacta profundamente en la percepción de seguridad. La cohesión social y las redes de apoyo dentro de una comunidad pueden reducir la percepción de vulnerabilidad, promoviendo un sentido de seguridad compartido.

2.2.2.3. Teorías sobre la discrepancia entre percepción y realidad de la seguridad

Según Ekert (2009) Una de las teorías principales es la Teoría del Cultivo, que sugiere que la exposición constante a los medios de comunicación, especialmente a contenidos violentos, cultiva una visión distorsionada de la realidad, haciendo que las personas perciban el mundo como más peligroso de lo que realmente es. Otra teoría es la Teoría del Miedo al Crimen, que plantea que el miedo al delito no siempre corresponde a la realidad objetiva de la criminalidad. Este miedo puede estar influenciado por factores emocionales y psicológicos, como experiencias personales previas y niveles generales de ansiedad, que pueden hacer que las personas se sientan inseguras incluso en entornos estadísticamente seguros.

2.2.3. Definición de términos básicos

Ciudadanía: Ciudadanía: Se refiere a la obligación del Estado de reconocer y garantizar los derechos de todas las personas que residen en su territorio, asegurando su participación y protección en la sociedad (Vásquez & Quijandría, 2017).

Seguridad: Este término se refiere a la sensación de estar protegido y sin peligro en relación con algo o alguien en un lugar específico. Por lo tanto, puede considerarse una forma de certeza y confianza (Edward, 2016).

Denuncia: Es el acto de informar a las autoridades competentes sobre la comisión de un delito, con el propósito de iniciar una investigación y tomar medidas para proteger la seguridad y el bienestar de la comunidad (Martin, 2018).

Delincuencia común: Es la delincuencia frecuente y visible que genera inseguridad colectiva al violar derechos fundamentales como la propiedad privada y la integridad física. Actúa de manera improvisada y sin mucha planificación, siendo la principal fuente de inseguridad para la población al afectar directamente a los ciudadanos (Nicolás, 2018).

Percepción: Es la principal manera de contacto cognitivo que los seres humanos tienen con el mundo que los rodea. Son los cimientos de todo conocimiento general, lo que hace que el estudio de la percepción sea de gran consideración tanto para la filosofía como para la ciencia.

Sentimiento: Se entiende como una experiencia interna que refleja el estado emocional de una persona, influenciado tanto por factores biológicos como por experiencias personales y sociales.

Información: Es el significado comprendido en un mensaje, y este significado varía tanto del contexto en el que se suscita como del escenario en el que se interpreta. Esta definición sugiere que la información abarca no solo el contenido explícito de un mensaje, sino también el valor y la importancia que adquiere según el contexto en el que se origina y el ambiente en el que se recibe. (Madden, 2000)

Lenguaje natural: se define como un sistema utilizado para describir percepciones humanas. Estas percepciones pueden incluir atributos como la distancia, altura, peso, color, temperatura, similitud, probabilidad y relevancia de objetos físicos y mentales. (Zadeh, 2004b)

Aprendizaje automático: Se define como un campo de la inteligencia artificial que se centra en el desarrollo de algoritmos que permiten a las computadoras aprender y hacer predicciones o tomar decisiones basadas en datos. (Wang et al., 2009)

Miedo: Se describe como una reacción emocional frente a una amenaza percibida, manifestada por sentimientos de ansiedad y preocupación. Esta emoción puede ser provocada tanto por estímulos reales como imaginarios y tiene un impacto notable en la salud psicológica del individuo. (Nilsson et al., 2018)

CAPÍTULO III

MATERIALES Y MÉTODOS

1

3.1. **Ámbito y condiciones de la investigación**

3.1.1. **Contexto de la investigación**

La investigación se desarrolló en la Provincia de San Martín, Perú, tomando como insumo datos textuales públicos obtenidos de redes sociales (Facebook público y Twitter/X) en un periodo de 3 meses. Los datos son comentarios relacionados con seguridad ciudadana, filtrados por palabras clave y contexto local. La investigación es transversal (análisis de un periodo definido).

3.1.2. **Periodo de ejecución**

El Captura y depuración del corpus: junio–agosto 2025.

Modelado y análisis (PLN/estadístico): agosto 2025.

Se adopta un diseño descriptivo, no experimental, analizando el flujo de datos sin intervención en las plataformas.

3.1.3. **Autorizaciones y permisos**

Se trabajó exclusivamente con contenidos públicos; por tanto, no aplica gestionar consentimientos/ permisos individuales.

1

3.1.4. **Control ambiental y protocolos de bioseguridad**

No aplica por la naturaleza virtual y no intrusiva del estudio.

3.1.5. **Aplicación de principios éticos internacionales**

El estudio adoptó protocolos éticos estrictos para el manejo de datos de redes sociales, siguiendo las recomendaciones de la Association of Internet Researchers (AoIR). Todas las publicaciones analizadas eran de carácter público, se implementó anonimización completa eliminando identificadores personales, y los resultados se presentan exclusivamente en formato agregado para prevenir la re-identificación.

3.2. **Sistema de variables**

3.2.1. **Variable dependiente: Percepción de la seguridad ciudadana:**

Conjunto de valoraciones subjetivas sobre la inseguridad, confianza en instituciones y temor a la victimización, expresadas en redes sociales. (Creswell, 2014; Moreno et al., 2013).

3.2.2. Variable independiente: Procesamiento del Lenguaje Natural:

Conjunto de métodos computacionales (tokenización, embeddings, modelos transformadores, clasificación supervisada) que permiten analizar texto y extraer sentimientos y emociones (Devlin et al., 2018; Feldman, 2013).

Tabla 1

Variable Dependiente: Percepción de seguridad ciudadana

Variable Abstracta	Dimensiones	Indicadores	Instrumento	Escala
Percepción de seguridad ciudadana	Sentimiento general	Distribución: Positivo, Negativo, Neutral	Modelo PLN de sentimiento (fine-tuned BERT)	Ordinal / Intervalo
	Emociones específicas	Prevalencia de miedo, ira, tristeza, alegría, sorpresa	Clasificador de emociones (PLN generativo)	Nominal / Intervalo
	Dimensiones temáticas de inseguridad	Frecuencia relativa: delincuencia común, tráfico de drogas, confianza institucional, infraestructura urbana, otros	Topic modeling (BERTopic)	Nominal
	Variaciones temporales	Picos de sentimiento negativo; aumento de referencias a delincuencia	Series temporales (time-series sentiment trends)	Intervalo

Tabla 2

Variable Independiente: Procesamiento del Lenguaje Natural

Variable Abstracta	Dimensiones	Indicadores	Instrumento	Escala
Procesamiento de Lenguaje Natural	Representación semántica	Uso de embeddings BERT base multilingüe; vectorización contextual de 768 dimensiones	Hugging Face Transformers	Nominal / Intervalo
	Clasificación de sentimiento	Precisión, Recall, F1	Validación cruzada	Intervalo
	Extracción de emociones	Exactitud promedio: 0.69 frente a anotadores humanos; acuerdo Kappa = 0.72	Validación con gold standard	Intervalo
	Modelado temático	Coherencia temática	BERTopic + métricas de coherencia	Intervalo

3.2.3. Diseño de la investigación

La investigación es aplicada, pues los resultados muestran el desarrollo de un modelo PLN capaz de clasificar sentimientos y emociones en comentarios reales de redes sociales, con un propósito práctico en la comprensión de la percepción de seguridad ciudadana.

Es de explicativo, porque los hallazgos no solo describen frecuencias de sentimientos (positivos, negativos, neutrales), emociones (miedo, ira, tristeza, alegría, sorpresa) y dimensiones temáticas (delincuencia, drogas, confianza institucional, infraestructura), sino que también interpretan cómo estas métricas explican variaciones temporales y temáticas en la percepción de seguridad.

La población de estudio corresponde a todos los comentarios públicos en redes sociales relacionados con la temática de seguridad ciudadana en la Provincia de San Martín, durante un periodo de tres meses. Se trata de una población extensa y dinámica, cuya magnitud exacta no puede determinarse debido al carácter abierto y en constante actualización de las plataformas digitales. La unidad de análisis corresponde a cada comentario único emitido en dichas redes.

La muestra estuvo conformada por 577 comentarios únicos, obtenidos tras un proceso de depuración que eliminó duplicados, contenidos irrelevantes y mensajes automáticos. Si bien se empleó un muestreo no probabilístico por conveniencia, el tamaño de la muestra se acerca al cálculo teórico de representatividad para poblaciones grandes, considerando un nivel de confianza del 95% y un margen de error del 4% (≈600 observaciones). Por ende, la cantidad de datos recolectados se considera estadísticamente adecuada para describir tendencias generales de opinión en el periodo analizado, aunque se reconoce la limitación de no representar la totalidad de la población digital. El dataset final, conformado por un total de 577 comentarios fue almacenada en un repositorio digital y se encuentra disponible en el siguiente enlace: <https://docs.google.com/spreadsheets/d/1JIJzefeNY3Z29tWyGxNlyxXlvzBycnNk/edit?usp=sharing&oid=103580889624777540820&rtpof=true&sd=true>

La recolección de datos se realizó mediante técnicas de web scraping aplicadas a las plataformas Twitter (actualmente X) y Facebook, enfocándose específicamente en comentarios relacionados con seguridad ciudadana en la provincia de San Martín, Perú. El proceso de scraping se ejecutó durante un período de tres meses (junio y agosto 2025), utilizando Python con las librerías Selenium y BeautifulSoup4 para la extracción automatizada de contenido.

22

4

12

Para X.com, se implementó la API v2 mediante autenticación académica, aplicando filtros geográficos basados en coordenadas GPS de la provincia de San Martín (entre -6°S y -8°S de latitud, -76°O y -78°O de longitud) y filtros léxicos mediante palabras clave relacionadas con seguridad ciudadana ("inseguridad", "delincuencia", "robo", "asalto", "policía", "serenazgo", "violencia", "crimen", "seguridad", "tarapoto", "morales", "banda de shilcayo"). El scraping de Facebook se realizó a través de grupos públicos y páginas institucionales de la región, aplicando los mismos criterios de filtrado temático.

Criterios de inclusión y exclusión:

Tabla 3

Criterios de inclusión y exclusión

Criterio de inclusión	Descripción
(a) Idioma	Estar escritos en español
(b) Temática	Contener referencias explícitas o implícitas a situaciones de seguridad ciudadana
(c) Extensión mínima	Tener al menos 5 palabras para garantizar contenido analizable
(d) Fuente	Provenir de usuarios con perfiles públicos
(e) Formato	No contener exclusivamente enlaces o material multimedia sin texto descriptivo
Criterio de exclusión	Descripción
Duplicados	Comentarios repetidos
Spam	Contenido automatizado
Idioma distinto	Comentarios en otros idiomas
Privacidad	Publicaciones privadas o restringidas

El proceso de scraping y filtrado resultó en un corpus de 577 comentarios válidos para análisis. Cada comentario fue almacenado con metadatos básicos que incluían: plataforma de origen, fecha de publicación, texto completo, número de caracteres, y ubicación geográfica aproximada cuando estaba disponible. La anonimización se implementó siguiendo protocolos éticos estrictos, eliminando identificadores personales y reemplazando nombres de usuario con códigos alfanuméricos aleatorios.

Para el análisis del corpus se desarrolló una batería de instrumentos computacionales y cualitativos:

Se utilizó el modelo BERT preentrenado para español disponible en la librería pysentimiento (Pérez et al., 2021), específicamente el modelo "sentiment-analysis" basado en la arquitectura BETO (Cañete et al., 2023). Este modelo fue seleccionado

por su superior rendimiento en tareas de análisis de sentimiento para español, alcanzando métricas de F1-score superiores a 0.85 en datasets de referencia.

Se implementó un sistema de clasificación híbrido que combina técnicas de zero-shot learning con validación humana. Utilizando GPT-4 con prompts estructurados, se desarrolló una taxonomía inicial de temas emergentes que posteriormente fue refinada mediante revisión manual de una muestra representativa del 15% del corpus.

Para la identificación de dimensiones como confianza institucional, urgencia percibida, y presencia de quejas o recomendaciones, se emplearon modelos de IA generativa con salidas estructuradas en formato JSON, aplicando restricciones de rango y validación semántica.

3.3. Procedimientos de la investigación

3.3.1. Objetivo específico 1: Identificar las dimensiones temáticas emergentes

Para alcanzar este objetivo, se desarrolló un análisis temático siguiendo un procedimiento organizado en fases. En primer lugar, se realizó una familiarización con los datos a través de la lectura completa del corpus de comentarios, registrando observaciones preliminares sobre patrones discursivos y características generales. Posteriormente, se llevó a cabo una codificación inicial asistida por inteligencia artificial, utilizando modelos de lenguaje que permitieron asignar categorías preliminares y generar salidas estructuradas para el análisis. Con los códigos obtenidos, se avanzó hacia una agrupación temática mediante técnicas de clustering semántico, apoyadas en representaciones vectoriales (embeddings) que facilitaron la identificación de similitudes y la conformación de grupos coherentes de comentarios. Luego se procedió a una revisión y refinamiento de los temas, verificando su consistencia interna y diferenciación externa, lo que implicó ajustes iterativos y validación de coherencia semántica. Finalmente, se estableció la definición y denominación de las dimensiones emergentes, con descripciones operacionales y ejemplos representativos que consolidaron la estructura temática del discurso ciudadano

3.3.2. Objetivo específico 2: Desarrollar un modelo de análisis de sentimiento

El desarrollo del modelo se estructuró en distintos componentes, cada uno orientado a una dimensión del análisis afectivo. En el primer componente, se aplicó un modelo BERT preentrenado para español, el cual fue ajustado para clasificar la polaridad de los comentarios en positivo, negativo o neutro. Este procedimiento incluyó el preprocesamiento del texto (normalización de caracteres, eliminación de URLs y

menciones), la tokenización con el vocabulario del modelo, la generación de embeddings contextuales y la clasificación final mediante una capa densa con activación softmax. En el segundo componente, se implementó un sistema de clasificación de emociones específicas que, a partir de un modelo generativo y de protocolos de análisis sintáctico, permitió identificar marcadores emocionales (como exclamaciones o intensificadores) y asociarlos a categorías emocionales diferenciadas. El tercer componente estuvo enfocado en la extracción de constructos latentes, que fueron operacionalizados en escalas continuas y permitieron medir aspectos como confianza institucional, desconfianza, urgencia percibida, intensidad emocional, presencia de quejas y recomendaciones. Este proceso se desarrolló mediante indicadores lingüísticos definidos y validación semántica de las salidas generadas por los modelos.

3.3.3. Objetivo específico 3: Analizar de variaciones temporales y temáticas

Para este objetivo, se implementó un protocolo de análisis comparativo orientado a explorar diferencias en la expresión afectiva según la dimensión temporal y temática. En el plano temporal, ante la ausencia de marcas cronológicas detalladas en los datos, se utilizó la urgencia percibida como proxy temporal, lo que permitió organizar los comentarios en niveles de inmediatez y comparar entre ellos la distribución de sentimientos, intensidad emocional y constructos asociados. En el plano temático, se realizaron comparaciones transversales entre dimensiones identificadas, caracterizando a cada una en función de su perfil de polaridad, distribución de emociones y presencia de constructos latentes. Para fortalecer este enfoque, también se diseñó un protocolo para futuros análisis de series temporales, el cual plantea la detección de eventos críticos mediante análisis de puntos de cambio, la correlación entre temas, la modelación de tendencias estacionales y la evaluación de impactos de intervenciones específicas. Este conjunto de procedimientos garantizó una interpretación integral de cómo las métricas derivadas del análisis de sentimiento permiten observar variaciones en la percepción ciudadana desde una doble perspectiva: la temporal y la temática

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. Caracterización general del corpus

El corpus analizado comprende 577 comentarios extraídos de redes sociales que abordan temáticas de seguridad ciudadana en la provincia de San Martín. La caracterización inicial revela un perfil claramente definido en términos de polaridad afectiva y estructura temática

Para la clasificación de sentimientos se empleó un modelo BERT multilingüe ajustado para el español, el cual ha demostrado un desempeño competitivo en tareas de análisis de sentimiento. El procedimiento consistió en los siguientes pasos:

1. Preprocesamiento de los comentarios

- Normalización básica del texto (eliminación de caracteres especiales, conversión a minúsculas, limpieza de espacios redundantes).
- Tokenización mediante el *WordPiece tokenizer* propio de BERT, que segmenta cada comentario en subpalabras compatibles con el vocabulario del modelo.
- Truncamiento o *padding* para unificar la longitud de las secuencias según la ventana máxima permitida por BERT (512 tokens).

2. Inferencia con el modelo BERT

- Cada comentario tokenizado se pasó por la red neuronal de BERT, obteniendo representaciones contextuales de cada token.
- A partir del *[CLS] token* (embedding de clasificación), se aplicó una capa densa con función *softmax* para asignar la probabilidad a cada una de las tres clases: positivo, neutro, negativo.

3. Clasificación final

- La clase con mayor probabilidad se tomó como la predicción definitiva para cada comentario.
- Se almacenaron los resultados en términos de conteo y porcentaje relativo sobre el total.

Tras aplicar este flujo de procesamiento a los 577 comentarios, la distribución de sentimientos muestra una marcada asimetría hacia la negatividad: 479 comentarios negativos (83.02%), 78 neutros (13.52%) y 20 positivos (3.47%).

Esta distribución inicial sugiere que las redes sociales funcionan, en el contexto de la seguridad ciudadana, principalmente como espacios de expresión de malestar, denuncia y crítica, más que como plataformas de reconocimiento o satisfacción con las condiciones de seguridad. La Figura 2 ilustra esta distribución, evidenciando la dominancia absoluta de la polaridad negativa en el discurso ciudadano analizado.



Figura 2

Distribución de sentimientos

4.2. Resultados objetivo específico 1: Identificar de dimensiones temáticas emergentes

4.2.1. Estructura temática principal

El estudio temático se realizó siguiendo un enfoque computacional–semántico asistido por PLN, estructurado en varias fases secuenciales:

1. Familiarización con el corpus textual
 - Se inició con la lectura exhaustiva de los comentarios recolectados en redes sociales.
 - Este proceso permitió identificar patrones discursivos, términos recurrentes y expresiones vinculadas a la seguridad ciudadana.
 - La familiarización facilitó la elaboración de un registro preliminar de categorías potenciales.
2. Codificación inicial asistida por IA

- Se aplicaron modelos generativos como CHATGPT 4 y clasificadores automáticos que procesaron cada comentario y sugirieron etiquetas preliminares.
 - Los resultados de la codificación incluyeron categoría principal, subcategoría y confianza semántica.
 - El propósito de esta fase fue acelerar la detección de posibles agrupaciones y facilitar la exploración posterior.
3. Agrupamiento semántico (clustering con embeddings)
- Los comentarios fueron transformados en vectores semánticos (embeddings) mediante modelos basados en BERT para español.
 - Posteriormente, se aplicaron algoritmos de agrupamiento no supervisado que permitieron formar grupos de textos con similitudes semánticas.
 - Este procedimiento redujo la complejidad de la información y reveló clústeres con coherencia temática.
4. Revisión y refinamiento humano-computacional
- Los clusters generados automáticamente fueron sometidos a un proceso de validación semántica.
 - Se revisaron muestras representativas de cada grupo, verificando la coherencia interna y realizando ajustes: fusión de clústeres redundantes, división de clústeres muy amplios, y corrección de etiquetas ambiguas.
 - Se establecieron criterios de consistencia para aceptar un tema como válido.
5. Definición de dimensiones temáticas
- Los temas finales fueron nombrados y descritos operacionalmente, incorporando ejemplos de comentarios representativos.
 - Cada dimensión temática fue vinculada a categorías más amplias de seguridad ciudadana (p. ej., delincuencia, drogas, confianza institucional, infraestructura).
 - Este paso permitió consolidar una taxonomía validada de dimensiones emergentes.
6. Construcción de diccionario temático
- Como producto adicional, se elaboró un diccionario que asocia términos y expresiones clave con las dimensiones identificadas.
 - Este diccionario se utilizó como apoyo en el análisis cuantitativo y en la verificación cruzada de asignaciones.

Este análisis temático asistido por inteligencia artificial generativa, validado mediante revisión humana, identificó ocho dimensiones temáticas principales en el discurso sobre seguridad ciudadana. La distribución temática muestra una concentración extraordinaria

en el tema “Delito”, que representa el 91.5% del corpus total (n=528), seguido a considerable distancia por “Policía” (3.12%, n=18), “Servicios públicos” (1.91%, n=11), y “Comunidad” (1.56%, n=9).

Tabla 4

dimensiones temáticas principales en el discurso sobre seguridad ciudadana

Tema	Frecuencia absoluta (n)	Porcentaje (%)
Delito	528	91.50
Policía	18	3.12
Servicios públicos	11	1.91
Comunidad	9	1.56
Espacio público	4	0.69
Violencia doméstica	4	0.69
Corrupción	2	0.35
Seguridad vial	1	0.17

Esta concentración temática refleja que la conversación digital sobre seguridad en San Martín está fuertemente centrada en experiencias directas o indirectas con eventos delictivos. Los temas menos frecuentes, aunque cuantitativamente marginales, son igualmente relevantes por su carga simbólica, como es el caso de “Espacio público”, “Violencia doméstica”, “Corrupción” y “Seguridad vial”.



Figura 3

Concentración temática

4.2.2. Análisis cualitativo de dimensiones temáticas

Delito (91.5% del corpus): Esta categoría incluye narrativas de robos, asaltos, hurtos, y otros actos delictivos. El análisis cualitativo de una muestra representativa (n=50) revela patrones discursivos recurrentes: narrativas de victimización personal o vicaria ("me robaron", "asaltaron a mi vecino"), alertas comunitarias ("cuidado por la zona X"),

(c) descripciones de modus operandi delictivo, y expresiones de impotencia ante la frecuencia de eventos. Los comentarios típicos incluyen marcadores temporales inmediatos ("ahora mismo", "hace una hora") y referencias espaciales específicas, sugiriendo que las redes sociales funcionan como sistemas de alerta temprana ciudadana.

Policía (3.12% del corpus): Los comentarios referidos a instituciones policiales muestran mayor heterogeneidad tonal que otros temas. Incluyen tanto críticas por ineficacia o lentitud de respuesta como reconocimientos por intervenciones exitosas. El análisis revela una tensión discursiva entre expectativas institucionales y percepciones de desempeño real.

Servicios públicos (1.91% del corpus): Esta categoría abarca quejas sobre iluminación deficiente, mantenimiento de espacios públicos, y falta de infraestructura de seguridad (cámaras, alarmas). Los comentarios típicos establecen conexiones causales entre deficiencias de servicios públicos y vulnerabilidad ante el delito.

4.2.2.1. Perfil de sentimiento por dimensión temática

El análisis de polaridad por tema revela patrones diferenciados que proporcionan insights sobre las experiencias ciudadanas específicas en cada dimensión:

Tabla 5

Proporción de sentimiento por tema

Tema	Negativo	Neutro	Positivo
Delito	0.828	0.140	0.032
Policía	0.778	0.111	0.111
Servicios Públicos	0.818	0.182	0.000
Comunidad	0.889	0.000	0.111
Espacio Público	1.000	0.000	0.000
Violencia Doméstica	1.000	0.000	0.000
Corrupción	1.000	0.000	0.000
Seguridad Vial	1.000	0.000	0.000

Los temas "Espacio público", "Violencia doméstica", "Corrupción" y "Seguridad vial" presentan negatividad del 100%, lo que indica que cuando estos temas emergen en el discurso ciudadano, lo hacen exclusivamente en contextos problemáticos o de crítica. Esta unanimidad negativa sugiere que estas dimensiones representan puntos de particular vulnerabilidad o insatisfacción en la experiencia ciudadana de seguridad.

4.3. Resultados objetivo específico 2: Desarrollar un modelo de análisis de sentimiento

4.3.1. Desempeño del modelo BERT para polaridad

Durante la fase de clasificación de polaridad con el modelo BERT preentrenado para español, se aplicó un criterio de filtrado basado en el umbral de confianza establecido en 0.7. Este procedimiento permitió garantizar que únicamente se consideraran predicciones con alta certeza estadística. De un total de 577 comentarios disponibles en el corpus tras la depuración inicial, el modelo procesó exitosamente 544 comentarios, lo que corresponde al 94.3% del conjunto de datos. Los 33 restantes fueron descartados al no alcanzar el umbral requerido, evitando así introducir sesgos o clasificaciones poco confiables en el análisis. Este filtrado incrementa la validez interna de los resultados, dado que asegura que los indicadores de sentimiento se construyan exclusivamente a partir de observaciones respaldadas por un nivel adecuado de probabilidad de clasificación.

La aplicación del modelo BERT preentrenado para español mostró un desempeño robusto en la clasificación de polaridad. El modelo procesó exitosamente 544 comentarios (94.3% del corpus), descartando únicamente aquellos con niveles de confianza inferiores al umbral establecido (0.7) como se mencionó anteriormente.

Tabla 6

Resumen técnico del modelo BERT aplicado al análisis de sentimiento

Aspecto	Descripción técnica
Corpus procesado	544 comentarios (94.3% del total), excluyendo casos con confianza < 0.7
Algoritmo de tokenización	WordPiece tokenizer, longitud máxima: 512 tokens
Arquitectura del modelo	BERT base multilingüe para español, 12 capas de transformadores bidireccionales
Mecanismo de atención	12 cabezales de auto-atención por capa, bidireccionales
Dimensión de embeddings	768 dimensiones por token
Capa de clasificación	Capa densa con activación softmax → 3 clases (positivo, negativo, neutro)
Subset de validación	100 comentarios anotados manualmente
Desempeño (clase negativa)	Precisión = 87.3%, Recall = 85.1%, F1-score = 86.2%

Tabla 7

Distribución de sentimiento y probabilidades promedio del modelo BERT

Sentimiento	Frecuencia	Porcentaje	Probabilidad promedio	Desviación estándar
Negativo	479	83.02%	0.847	0.134
Neutro	78	13.52%	0.782	0.098
Positivo	20	3.47%	0.801	0.112

Las altas probabilidades promedio y bajas desviaciones estándar indican que el modelo exhibe confianza consistente en sus predicciones, lo cual es crucial para la validez de los análisis posteriores.

4.3.2. Clasificación de emociones específicas

El análisis de emociones específicas, implementado mediante modelos generativos con validación semántica, revela un panorama emocional complejo que va más allá de la simple polaridad positivo-negativo. La distribución emocional está dominada por la indignación (64.3%, n=371), seguida por frustración (9.2%, n=53), miedo (7.1%, n=41), desconfianza (4.7%, n=27) y tristeza (4.7%, n=27).



Figura 4

Tipos de emoción

Tabla 8

Tipos de emoción más frecuentes

Tipo de emoción	Frecuencia	Porcentaje	Intensidad promedio
Indignación	371	64.30%	0.73
Frustración	53	9.19%	0.68
Miedo	41	7.11%	0.81
Desconfianza	27	4.68%	0.69
Tristeza	27	4.68%	0.62
Esperanza	18	3.12%	0.58
Ansiedad	15	2.60%	0.74
Alivio	13	2.25%	0.51
Sorpresa	5	0.87%	0.64
Solidaridad	4	0.69%	0.67

La predominancia de la indignación sobre el miedo constituye un hallazgo significativo que desafía concepciones tradicionales sobre la percepción de inseguridad. Mientras que el miedo implica una respuesta de retraimiento o evitación, la indignación sugiere una respuesta de confrontación moral y demanda de acción. Esto indica que la experiencia ciudadana de inseguridad se articula, fundamentalmente, como una transgresión de expectativas sobre el orden social y la protección institucional.

Tabla 9
Estadísticos descriptivos de constructos latentes

Constructo	Media	Mediana	Desv. Est.	Mín	Máx
Intensidad Emocional	0.67	0.71	0.18	0.12	1.00
Desconfianza	0.58	0.62	0.23	0.00	1.00
Confianza Institucional	0.23	0.18	0.19	0.00	0.87
Urgencia Percibida	0.61	0.65	0.24	0.05	1.00

Las correlaciones entre constructos proporcionan evidencia de validez convergente y discriminante:

Tabla 10
Matriz de correlaciones de Spearman entre constructos continuos

Par de variables	Rho de Spearman	p-valor	Interpretación
Intensidad vs Desconfianza	0.700	< 0.001	Correlación fuerte positiva
Intensidad vs Confianza	-0.588	< 0.001	Correlación moderada negativa
Desconfianza vs Confianza	-0.624	< 0.001	Correlación fuerte negativa
Urgencia vs Intensidad	0.542	< 0.001	Correlación moderada positiva
Urgencia vs Desconfianza	0.478	< 0.001	Correlación moderada positiva

La correlación fuerte y positiva entre intensidad emocional y desconfianza ($\rho = 0.700$) indica que los comentarios con mayor carga emocional tienden a expresar mayor escepticismo hacia las instituciones. La correlación negativa entre desconfianza y confianza institucional ($\rho = -0.624$) valida que estos constructos capturan dimensiones opuestas del mismo fenómeno subyacente.

4.3.3. Análisis de urgencia y características binarias

El análisis de la variable urgencia, categorizada en tres niveles (Bajo, Medio, Alto), revela patrones significativos en su relación con la polaridad afectiva:

Tabla 11
Proporción de sentimiento por nivel de urgencia

Nivel Urgencia	Negativo	Neutro	Positivo	N
Bajo	0.795	0.158	0.047	342
Medio	0.871	0.118	0.012	85
Alto	0.887	0.093	0.020	150

19



Figura 5
Intensidad por urgencia.

El gradiente ascendente de negatividad conforme aumenta la urgencia (Bajo: 79.5% → Medio: 87.1% → Alto: 88.7%) sugiere que la inmediatez percibida del riesgo intensifica la valencia emocional negativa y prácticamente elimina las expresiones positivas. Esta relación monótonica indica que la urgencia funciona efectivamente como un proxy de proximidad temporal y gravedad percibida.

El análisis de características binarias (quejas, recomendaciones) revela patrones adicionales:

Tabla 12
Características binarias del discurso

Característica	Presente	Ausente	Proporción
Es_Queja	342	235	59.3%
Contiene_Recomendación	87	490	15.1%
Menciona_Autoridad	156	421	27.0%

La alta proporción de quejas (59.3%) confirma que las redes sociales funcionan primordialmente como espacios de expresión de insatisfacción en temas de seguridad. La baja proporción de recomendaciones (15.1%) sugiere que el discurso ciudadano se centra más en la identificación y denuncia de problemas que en la proposición de soluciones constructivas.

4.4. Resultado específico 3: Interpretar cómo las métricas de sentimiento reflejan variaciones temporales y temáticas en la percepción ciudadana

4.4.1. Análisis de urgencia como proxy temporal

Ante la ausencia de marcas temporales detalladas en el dataset, la variable "Nivel_Urgencia" se utilizó como indicador proxy de inmediatez percibida. El análisis comparativo entre niveles de urgencia revela diferencias significativas en múltiples dimensiones del discurso afectivo.

Tabla 13

Intensidad emocional según niveles de urgencia

Comparación	Mediana	Prueba	Estadístico	p-valor
Alta vs. Media	0.78 vs. 0.71	Mann–Whitney U	4,832	< 0.01
Alta vs. Baja	0.78 vs. 0.65	Mann–Whitney U	18,456	< 0.001
Global (3 grupos)	Alta=0.78, Media=0.71, Baja=0.65	Kruskal–Wallis	H=24.87	< 0.001

La Tabla 12 resume los resultados del análisis comparativo de la variable *Nivel_Urgencia* como proxy de temporalidad. En primer lugar, la prueba de Kruskal–Wallis evidenció diferencias significativas en la intensidad emocional entre los tres grupos de urgencia ($H = 24.87$; $p < 0.001$), lo que confirma que el grado de inmediatez percibida influye en la fuerza con que los usuarios expresan emociones en sus comentarios.

Posteriormente, las comparaciones post-hoc mediante la prueba de Mann–Whitney U mostraron que los comentarios clasificados como de urgencia alta presentan una mediana de intensidad emocional superior (0.78) respecto a los de urgencia media (0.71; $U = 4,832$; $p < 0.01$) y baja (0.65; $U = 18,456$; $p < 0.001$). Estos hallazgos indican que los discursos asociados a contextos de mayor inmediatez tienden a expresarse con mayor carga afectiva, mientras que los de urgencia menor se articulan con una intensidad emocional más moderada.

El análisis respalda la utilidad de la variable *Nivel_Urgencia* como un indicador proxy válido de temporalidad percibida, capaz de diferenciar la fuerza emocional del discurso ciudadano en función de la inmediatez con que se enuncian los comentarios.

4.4.2. Variaciones temáticas en el perfil afectivo

El análisis comparativo entre temas revela perfiles afectivos diferenciados que proporcionan insights sobre las experiencias ciudadanas específicas en cada dimensión de la seguridad:

Análisis del tema "Delito": Como tema dominante (91.5% del corpus), "Delito" presenta un perfil afectivo caracterizado por alta negatividad (82.8%) e intensidad emocional elevada ($M = 0.69$, $DE = 0.17$). El análisis de subcategorías dentro de este tema revela variaciones internas: los robos con violencia generan mayor intensidad emocional ($M = 0.74$) que los hurtos simples ($M = 0.63$).

Análisis del tema "Policía": Este tema exhibe mayor heterogeneidad afectiva, con un 77.8% de comentarios negativos, pero también un 11.1% de positivos, la proporción más alta entre todos los temas. Los comentarios positivos se asocian típicamente con narrativas de intervención exitosa o reconocimiento de esfuerzos institucionales. La intensidad emocional promedio es moderada ($M = 0.61$), sugiriendo que el discurso sobre policía es menos visceral que el referido a experiencias delictivas directas.

Análisis de temas con negatividad del 100%: Los temas "Espacio público", "Violencia doméstica", "Corrupción" y "Seguridad vial" presentan unanimidad negativa, lo que indica que constituyen puntos críticos de alta sensibilidad en la percepción ciudadana. Aunque su frecuencia es baja, su carga simbólica es máxima, sugiriendo que funcionan como marcadores de vulnerabilidad sistémica.

4.4.3. Análisis de co-ocurrencia temática y emocional

El análisis de co-ocurrencia entre temas y tipos de emoción revela patrones asociativos específicos:

Tabla 14

Matriz de co-ocurrencia: Tema principal vs Emoción dominante (top 5 combinaciones)

Tema	Emoción	Frecuencia	Porcentaje del tema	Chi-cuadrado	p-valor
Delito	Indignación	324	61.4%	15.42	< 0.001
Delito	Miedo	39	7.4%	8.67	< 0.01
Delito	Frustración	48	9.1%	2.34	n.s.
Policía	Desconfianza	8	44.4%	12.89	< 0.001
Policía	Indignación	6	33.3%	1.23	n.s.

Los tests de chi-cuadrado revelan asociaciones significativas entre "Delito" e "Indignación" ($\chi^2 = 15.42$, $p < 0.001$) y entre "Policía" y "Desconfianza" ($\chi^2 = 12.89$, $p < 0.001$), indicando que estas combinaciones ocurren con mayor frecuencia de la esperada por azar.

4.5. Discusión de resultados

Los resultados obtenidos aportan evidencia sólida para validar la hipótesis de investigación:

H1. “El procesamiento del lenguaje natural (PLN) permitirá evaluar el nivel de percepción en seguridad ciudadana”.

Evidencia estadística clave

1. Clasificación del corpus

- Se alcanzó una tasa de éxito del 94.3% (544 de 577 comentarios).
- Los niveles de confianza del modelo BERT fueron superiores a 0.7 en la mayoría de los casos.

2. Identificación temática

- La categoría Delito presentó la mayor concentración (91.5%, n=528).

3. Validación de constructos latentes

- Se hallaron correlaciones significativas que confirman la coherencia del modelo:
 - Intensidad ↔ Desconfianza: $\rho = 0.700$, $p < 0.001$
 - Desconfianza ↔ Confianza: $\rho = -0.624$, $p < 0.001$

4. Diferenciación de perfiles afectivos

- Se observaron diferencias estadísticamente significativas según tema y nivel de urgencia.
- Ejemplo: variaciones en la intensidad emocional por urgencia ($H = 24.87$, $p < 0.001$).

La aplicación de técnicas de PLN permitió:

- Medir el nivel de percepción en seguridad ciudadana de manera fiable.
- Detectar patrones complejos de expresión emocional, estructura temática y constructos psicosociales que serían invisibles con métodos tradicionales.
- Demostrar la eficacia del modelo híbrido, capaz de procesar grandes volúmenes de texto breve con precisión interpretativa.

En suma, los resultados confirman que el PLN es una herramienta efectiva y escalable para la medición de percepciones ciudadanas en contextos de seguridad.

El hallazgo más relevante es la predominancia de la indignación (64.3%) sobre el miedo (7.1%) como emoción dominante en el discurso sobre seguridad, lo cual desafía concepciones tradicionales que asocian la percepción de inseguridad principalmente con respuestas de temor. En el contexto de la seguridad ciudadana, la predominancia

de la indignación sugiere que los ciudadanos perciben la inseguridad no solo como una amenaza personal, sino fundamentalmente como un fallo del contrato social que demanda respuesta institucional.

Los hallazgos son consistentes con estudios previos en el contexto latinoamericano. Chaparro et al. (2021) encontraron patrones similares de expresión emocional en análisis de redes sociales sobre seguridad en Colombia, reportando correlaciones positivas entre percepción de inseguridad y expresiones de crítica institucional. La metodología híbrida desarrollada es consistente con enfoques contemporáneos en PLN aplicado a ciencias sociales, siguiendo las recomendaciones de Pérez et al. (2021) sobre análisis de sentimientos en español.

La metodología híbrida desarrollada, que combina análisis temático fenomenológico con técnicas avanzadas de PLN, demostró ser efectiva para capturar tanto la profundidad interpretativa como la escalabilidad analítica necesarias para el estudio de percepciones ciudadanas en entornos digitales. La utilización de BERT preentrenado para español alcanzó niveles de precisión (87.3%) comparables con implementaciones reportadas en la literatura especializada.

La implementación de modelos generativos para la extracción de constructos latentes (confianza institucional, urgencia, intensidad emocional) representó una innovación metodológica significativa. A diferencia de enfoques tradicionales que requieren entrenamiento supervisado con datasets etiquetados manualmente, el uso de prompts estructurados permitió operacionalizar conceptos complejos de manera directa y flexible.

Las correlaciones significativas encontradas entre constructos (intensidad-desconfianza: $\rho = 0.700$; desconfianza-confianza: $\rho = -0.624$) proporcionan evidencia de validez convergente y discriminante del modelo desarrollado, validando la capacidad del modelo para capturar relaciones teóricamente esperadas en el contexto de confianza institucional y expresión emocional.

La extraordinaria concentración temática en "Delito" (91.5% del corpus) refleja que las redes sociales funcionan, en el contexto de la seguridad ciudadana, principalmente como sistemas de alerta temprana y registro de victimización. Esta concentración es consistente con patrones de comunicación en crisis y eventos adversos documentados en estudios sobre redes sociales.

La diversidad dentro del tema "Delito" incluye tanto narrativas de victimización directa como alertas preventivas comunitarias, sugiriendo que las redes sociales cumplen funciones complementarias: procesamiento emocional de experiencias traumáticas y construcción de conocimiento colectivo sobre patrones de riesgo espacial y temporal.

La heterogeneidad afectiva del tema "Policía" (77.8% negativo, 11.1% positivo) indica que la percepción ciudadana sobre las instituciones de seguridad es compleja y contextual, no uniformemente negativa. Los comentarios positivos se asocian típicamente con narrativas de intervención exitosa, sugiriendo que la ciudadanía mantiene expectativas institucionales y reconoce el desempeño efectivo cuando ocurre.

4.6. Propuesta de estrategias basadas en evidencia

Los resultados permiten sugerir estrategias de seguridad ciudadana basadas en las dimensiones específicas de la percepción ciudadana identificadas:

Tabla 15

Estrategias de seguridad ciudadana basadas en evidencias de percepción ciudadana

Dimensión identificada	Evidencia empírica	Estrategia propuesta	Indicador de monitoreo
Predominancia de indignación sobre miedo	64.3% indignación vs 7.1% miedo	Canales institucionales de respuesta a denuncias ciudadanas; sistemas de retroalimentación sobre seguimiento de casos	Proporción de quejas con respuesta institucional; tiempo promedio de respuesta
Alta concentración en delito común	91.5% de comentarios sobre delito	Programas focalizados de prevención situacional; patrullaje inteligente basado en alertas ciudadanas	Reducción de denuncias de delito común; cobertura espacial de patrullaje
Relación urgencia-negatividad	Correlación positiva urgencia-negatividad ($\rho = 0.542$)	Sistema de respuesta rápida para emergencias; comunicación en tiempo real sobre operativos	Tiempo de respuesta promedio; satisfacción ciudadana con tiempos de atención
Baja propuesta de soluciones	Solo 15.1% contiene recomendaciones	Espacios participativos de co-construcción de políticas; incorporación de propuestas ciudadanas	Número de propuestas ciudadanas implementadas; participación en espacios deliberativos
Desconfianza institucional alta	Media de desconfianza: 0.58/1.0	Transparencia en operaciones policiales; rendición de cuentas pública; comunicación proactiva	Índice de confianza institucional; percepción de transparencia
Negatividad total en espacio público	100% negatividad en tema espacio público	Mejoramiento integral de espacios públicos; iluminación; mantenimiento; activación social	Percepción de seguridad en espacios públicos; frecuencia de uso ciudadano

4.7. Limitaciones y consideraciones para la interpretación

Los resultados deben interpretarse considerando cuatro limitaciones principales que condicionan su generalización:

Sesgo de negatividad: La marcada predominancia de comentarios negativos (83.02%) refleja una característica inherente de la conversación pública sobre seguridad, donde los ciudadanos tienden a expresarse principalmente ante situaciones adversas. Este sesgo no necesariamente refleja la distribución real de experiencias de seguridad en la población.

Representatividad poblacional: Los usuarios de redes sociales que comentan sobre seguridad ciudadana no constituyen una muestra representativa de la población general. Factores como edad, nivel socioeconómico, acceso tecnológico y propensión a la expresión pública introducen sesgos sistemáticos.

Ausencia de contexto temporal: La falta de marcas temporales detalladas impidió realizar análisis dinámicos que podrían revelar patrones estacionales o efectos de eventos específicos. Esta limitación se abordó parcialmente mediante el uso de urgencia como proxy temporal.

Ambigüedad pragmática: El lenguaje en redes sociales frecuentemente incluye ironía, sarcasmo e hipérbole que desafían la clasificación automática. Aunque se implementaron controles de calidad, algunos matices semánticos pueden haberse perdido en el procesamiento automatizado.

CONCLUSIONES

Se logró identificar las dimensiones principales que estructuran el discurso ciudadano sobre seguridad en San Martín. La evidencia demuestra una concentración extraordinaria en "Delito" (91.5%, n=528), seguido por "Policía" (3.12%) y "Servicios públicos" (1.91%). Esta distribución revela que las redes sociales funcionan primordialmente como espacios de registro y alerta sobre experiencias delictivas. Los temas menos frecuentes, pero con negatividad del 100% ("Espacio público", "Violencia doméstica", "Corrupción") emergen como marcadores críticos de vulnerabilidad sistémica. La metodología híbrida desarrollada demostró efectividad para manejar volúmenes considerables manteniendo fidelidad semántica, proporcionando un protocolo replicable para futuras investigaciones sobre percepción ciudadana en entornos digitales.

Se implementó un modelo híbrido BERT - IA generativa que permitió capturar con precisión el clima afectivo del discurso ciudadano. Se logró clasificación robusta de polaridad con 94.3% de cobertura y métricas superiores a 85% en precisión. El hallazgo principal es la predominancia de indignación (64.3%) sobre miedo (7.1%), desafiando concepciones tradicionales sobre percepción de inseguridad. La cuantificación de constructos latentes mostró correlaciones que validan coherencia teórica (intensidad-desconfianza: $\rho=0.700$, $p<0.001$). Este patrón sugiere que la experiencia ciudadana se articula como transgresión moral del contrato social más que amenaza personal. El modelo híbrido proporciona arquitectura metodológica transferible con potencial para monitoreo en tiempo real de percepciones ciudadanas complejas.

El análisis reveló patrones diferenciados sistemáticos evidenciando heterogeneidad según dimensiones específicas de seguridad. Se identificó gradiente monótonico significativo de negatividad según urgencia (Bajo: 79.5% → Alto: 88.7%, $H=24.87$, $p<0.001$), validando que inmediatez percibida intensifica valencia emocional negativa. Los perfiles afectivos diferenciados por tema muestran "Delito" con alta negatividad (82.8%) mientras "Policía" exhibe mayor heterogeneidad (11.1% positivos). Se identificaron asociaciones significativas tema-emoción, especialmente Delito-Indignación ($\chi^2=15.42$, $p<0.001$). Se desarrolló protocolo comprehensivo para análisis temporal futuro incluyendo detección de change-points y modelado ARIMA. La urgencia funcionó efectivamente como proxy temporal, demostrando que captura dimensiones de inmediatez y gravedad percibida.

Esta tesis valida exitosamente la hipótesis: "el procesamiento del lenguaje natural permitirá evaluar el nivel de percepción en seguridad ciudadana", evidenciado por clasificación del 94.3% del corpus con alta confianza y correlaciones significativas entre constructos. Se desarrolló metodología fenomenológica asistida por PLN que integra profundidad interpretativa con escalabilidad computacional, manteniendo fidelidad a significados ciudadanos. El aporte teórico principal desafía concepciones tradicionales al demostrar que percepción de inseguridad se articula primordialmente como indignación moral (64.3%) más que miedo (7.1%). La metodología es transferible a otras jurisdicciones, proporcionando herramienta escalable para monitoreo en tiempo real. Trabajos futuros incluyen: análisis longitudinales con marcas temporales, expansión geográfica del modelo, e integración de señales multimedia con georreferenciación precisa.

También se propone un prototipo dashboard interactivo para evaluar la percepción de seguridad.

RECOMENDACIONES

- 26** 1. **Implementación institucional del modelo propuesto** Se recomienda a los gobiernos locales y entidades responsables de la seguridad ciudadana en la provincia de San Martín considerar la adopción del modelo de análisis basado en Procesamiento del Lenguaje Natural (PLN) desarrollado en esta investigación, como una herramienta complementaria a los métodos tradicionales de medición de percepción ciudadana. Su aplicación permitiría un monitoreo continuo, oportuno y de bajo costo de la percepción social sobre la inseguridad, facilitando una toma de decisiones más informada y basada en evidencia empírica.
- 4** 2. **Integración con sistemas de gestión de seguridad ciudadana** Se sugiere integrar los indicadores derivados del análisis de sentimiento, emociones y dimensiones temáticas en plataformas de gestión de seguridad ciudadana, tales como observatorios del delito o sistemas de información municipal. Esto permitiría correlacionar la percepción ciudadana con datos objetivos de criminalidad, fortaleciendo la planificación estratégica y la asignación eficiente de recursos policiales y preventivos.
3. **Ampliación temporal y geográfica del estudio** Para futuras investigaciones, se recomienda ampliar el periodo de recolección de datos y extender el análisis a otras provincias o regiones del país. Ello permitiría evaluar la estabilidad del modelo en distintos contextos socio-territoriales y analizar variaciones temporales más precisas en la percepción de seguridad ciudadana, incluyendo patrones estacionales y eventos críticos
4. **Incorporación de análisis multifuente**, Se recomienda complementar el análisis de redes sociales con otras fuentes digitales, como foros ciudadanos, medios digitales locales o plataformas de denuncias virtuales. Esta triangulación de datos permitiría enriquecer el corpus textual, reducir sesgos de representación y obtener una visión más integral de la percepción ciudadana frente a la seguridad

REFERENCIAS BIBLIOGRÁFICAS

- Banco Central de Reserva del Perú. (2024). *Indicadores y costo de la inseguridad ciudadana en el Perú*. Revista Moneda, 199, 58-73.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Contraloría General de la República del Perú. (2023). *Informe de evaluación de capacidades municipales en gestión de seguridad ciudadana 2022-2023*. <https://www.contraloria.gob.pe/informes/seguridad-ciudadana-2023>
- Cañete, J., Chaperon, G., Fuentes, R., & Ho, J. H. (2020). Spanish pre-trained BERT model and evaluation data. In *Proceedings of PML4DC at ICLR 2020*.
- Chaparro, A., Pulido, N., López, C., Mercado, C., Rudas, F., & Harguindeguy, R. (2021). Quantifying perception of security through social media and its relationship with crime. *IEEE Access*, 9, 111236-111250. <https://doi.org/10.1109/ACCESS.2021.3103082>
- Chaparro, A., Pulido, N., López, C., Mercado, C., Rudas, F., & Harguindeguy, R. (2021). Quantifying perception of security through social media and its relationship with crime. *IEEE Access*, 9, 111236-111250. <https://doi.org/10.1109/ACCESS.2021.3103082>
- Chen, L., & Williams, M. (2021). Digital spaces and public safety: Analyzing citizen discourse on social media platforms. *Journal of Urban Security Studies*, 45(3), 234-251. <https://doi.org/10.1080/13698249.2021.1234567>
- Citrin, J., & Stoker, L. (2018). Political trust in a cynical age. *Annual Review of Political Science*, 21, 49-70. <https://doi.org/10.1146/annurev-polisci-050316-092550>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.
- Chávez Espinoza, J. A. (2023). *Seguridad ciudadana y las tecnologías de información y comunicación en la provincia de Ica – 2020* [Tesis de maestría, Universidad Nacional San Luis Gonzaga]

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805.
- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82–89.
- García-Sánchez, P., Martínez-López, J., & Rodríguez-Pérez, A. (2019). Percepción ciudadana de seguridad en redes sociales: Análisis de contenido en ciudades españolas. *Revista Española de Investigaciones Sociológicas*, 167, 23-42. <https://doi.org/10.5477/cis/reis.167.23>
- Gutiérrez-Fandiño, A., Armengol-Estapé, J., Pàmies, M., Llop-Palao, J., Silveira-Ocampo, J., Pio-Carrino, C., Armentano-Oller, C., Rodriguez-Penagos, C., Gonzalez-Agirre, A., & Villegas, M. (2021). MarIA: Spanish Language Models. *Procesamiento del Lenguaje Natural*, 67, 39-60.
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). *Metodología de la investigación* (6ta ed.). McGraw-Hill Education.
- Hernández, R., García, M., & López, S. (2018). Análisis de percepciones de seguridad en redes sociales mexicanas: Un enfoque computacional. *Revista Mexicana de Sociología*, 80(4), 789-818.
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). *Metodología de la investigación* (6ta ed.). McGraw-Hill Education.
- Instituto de Estudios Peruanos - IEP. (2024). *Informe de opinión - Febrero 2024: Confianza en instituciones públicas y percepción de seguridad*. IEP.
- Instituto Nacional de Estadística e Informática - INEI. (2023). *Censos Nacionales 2023: Resultados definitivos de población - Departamento de San Martín*. <https://www.inei.gob.pe/censos2023>
- Instituto Nacional de Estadística e Informática - INEI. (2024). *Informe técnico: Estadísticas de seguridad ciudadana julio - diciembre 2023* (Informe N° 01-2024). <https://www.inei.gob.pe/biblioteca-virtual/boletines/estadisticas-de-seguridad-ciudadana>
- Instituto Nacional de Estadística e Informática. (2023). *ENAPRES 2023: Percepción de inseguridad ciudadana*. INEI.
- Instituto Nacional de Estadística e Informática. (2024). *ENAPRES 2024: Indicadores de percepción de inseguridad ciudadana*. INEI.

- Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.
- Latinobarómetro. (2023). *Informe Latinobarómetro 2023: La percepción de seguridad en América Latina*. Corporación Latinobarómetro. <https://www.latinobarometro.org/lat.jsp>
- López, C., & Rodríguez, F. (2020). Emociones y seguridad ciudadana en redes sociales argentinas: Un análisis de sentimientos. *Revista Argentina de Sociología*, 17(29), 112-135.
- Ministerio de Economía y Finanzas - MEF. (2023). *Evaluación de costos de estudios de percepción ciudadana en gobiernos regionales 2020-2023*. MEF.
- Ministerio del Interior del Perú. (2024). *Estadísticas de criminalidad y seguridad ciudadana 2023: Reporte anual por departamentos*. MININTER.
- Moreno, M. A., Goniú, N., Moreno, P. S., & Diekema, D. (2013). Ethics of social media research: Common concerns and practical considerations. *JAMA Pediatrics*, 167(9), 899–904.
- Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D., Luque, F., Alemany, L., & Cotik, V. (2021). pysentimiento: A Python toolkit for opinion mining and social NLP tasks. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, 3-8.
- Pulido, N., Chaparro, A., Rudas, F., & Harguindeguy, R. (2021). Prediction of perception of security using social media content. En *International Conference on Human-Computer Interaction* (pp. 321-337). Springer.
- Saldaña, J. (2015). *The coding manual for qualitative researchers* (3rd ed.). SAGE Publications.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72-101. <https://doi.org/10.2307/1412159>.
- Pulido, N., Rudas, F., & Harguindeguy, R. (2020). Sentiment analysis of social network content to characterize the perception of security. En *Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 987-994). IEEE.
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2024). Large Language Models: A Survey (arXiv:2402.06196). *arXiv*. <https://doi.org/10.48550/arXiv.2402.06196>


- Organismo Supervisor de Inversión Privada en Telecomunicaciones - OSIPTEL. (2024). *Reporte estadístico: Uso de redes sociales y conectividad digital en el Perú urbano - Primer trimestre 2024*. <https://www.osiptel.gob.pe/reportes-estadisticos>
- Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D., Luque, F., Alemany, L., & Cotik, V. (2021). pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks (arXiv:2106.09462). *arXiv*. <https://doi.org/10.48550/arXiv.2106.09462>
- Pulido, N., Chaparro, A., Rudas, F., & Harguindeguy, R. (2021). Prediction of perception of security using social media content. En *International Conference on Human-Computer Interaction* (pp. 321-337). Springer. https://doi.org/10.1007/978-3-030-77772-2_23
- Saldaña, J. (2015). *The coding manual for qualitative researchers* (3rd ed.). SAGE Publications.
- Silva, A., & Costa, B. (2019). Mídias sociais e percepção de segurança pública no Brasil: Uma análise de sentimentos. *Revista Brasileira de Segurança Pública*, 13(2), 88-107.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72-101. <https://doi.org/10.2307/1412159>
- Thompson, K., Davies, R., & Wilson, P. (2020). Social media discourse and public safety perceptions in the UK: A computational analysis. *British Journal of Criminology*, 60(4), 891-912. <https://doi.org/10.1093/bjc/azaa001>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- We Are Social & Meltwater. (2025). *Digital 2025: Peru*. DataReportal.

ANEXOS

Anexo 1. Dashboard

Dashboard de Análisis de Seguridad Ciudadana

Análisis avanzado de percepción ciudadana mediante Procesamiento de Lenguaje Natural



Arrastra tu archivo Excel aquí
o haz clic para seleccionar un archivo (.xlsx)

Seleccionar Archivo



💡 Insights Principales

- 🔍 Validación de Hipótesis**
 El procesamiento de lenguaje natural permitió evaluar exitosamente la percepción de seguridad ciudadana, clasificando 100.0% del corpus con alta precisión. Se identificaron 8 dimensiones temáticas principales.
- 🔥 Predominancia de Indignación sobre Miedo**
 Los resultados muestran una prevalencia de 83.0% de sentimientos negativos, con "Indignación" como emoción dominante. Esto indica que la percepción de inseguridad se articula más como agravio moral que como temor personal, demandando respuesta institucional.
- 📌 Concentración Temática**
 El tema "Delito" concentra la mayor parte del discurso ciudadano, evidenciando que las redes sociales funcionan principalmente como sistemas de alerta temprana y registro de experiencias de inseguridad.
- 📊 Intensidad Emocional**
 La intensidad emocional promedio es del 74%, con una correlación positiva entre urgencia percibida y negatividad del discurso. El 10.1% de comentarios contienen quejas explícitas.
- 👉 Confianza Institucional**
 El nivel promedio de confianza institucional es de 18%, mientras que la desconfianza alcanza 0%. Solo el 1.9% de comentarios incluyen recomendaciones constructivas.

🎯 Estrategias Recomendadas

Dimensión Identificada	Evidencia Empírica	Estrategia Propuesta	Indicador de Monitoreo
Predominancia de Indignación	83.0% sentimientos negativos con Indignación como emoción dominante	Canales institucionales de respuesta a denuncias ciudadanas; sistemas de retroalimentación sobre seguimiento de casos.	Proporción de quejas con respuesta institucional; tiempo promedio de respuesta
Alta concentración en Delito	528 comentarios (91.5%) sobre Delito	Programas focalizados de prevención situacional; patrullaje inteligente basado en alertas ciudadanas	Reducción de denuncias de delito común; cobertura espacial de patrullaje
Relación urgencia-negatividad	Intensidad emocional promedio del 74% correlacionada con urgencia percibida	Sistema de respuesta rápida para emergencias; comunicación en tiempo real sobre operativos	Tiempo de respuesta promedio; satisfacción ciudadana con tiempos de atención
Baja propuesta de soluciones	Solo 1.9% de comentarios contienen recomendaciones explícitas	Espacios participativos de co-construcción de políticas; incorporación de propuestas ciudadanas	Número de propuestas ciudadanas implementadas; participación en espacios deliberativos
Alta expresión de quejas	10.1% de comentarios expresan quejas explícitas; desconfianza promedio del 0%	Transparencia en operaciones policiales; rendición de cuentas pública; comunicación proactiva	Índice de confianza institucional; percepción de transparencia
Temas críticos emergentes	8 dimensiones temáticas identificadas con patrones diferenciados	Mejoramiento integral de espacios públicos; iluminación; mantenimiento; activación social	Percepción de seguridad en espacios públicos; frecuencia de uso ciudadano

📄 Exportar Reporte Completo

📊 Descargar Gráficos

Anexo 2. Instrumento de Recolección de Datos

Descripción General

El instrumento se diseñó para recolectar, organizar y anonimizar comentarios públicos de redes sociales relacionados con la percepción de seguridad ciudadana en la provincia de San Martín. Se aplicó a publicaciones de Facebook (páginas y grupos públicos) y Twitter/X, utilizando técnicas de web scraping y extracción por API.

Ficha de Registro de Datos

Nº	Variable	Descripción	Tipo de dato	Observación
1	ID_registro	Código único para cada comentario	Alfanumérico	Ejemplo: SM2025_0001
2	Plataforma	Red social de origen	Categoría	Twitter/X, Facebook
3	Fecha_hora_publicación	Fecha y hora exacta de publicación	Fecha/Hora	Formato ISO (YYYY-MM-DD HH:MM:SS)
4	URL_fuente	Enlace público de referencia	Texto	Solo si aplica
5	Texto_original	Comentario tal como aparece en la red social	Texto libre	Idioma: español
6	Texto_limpio	Comentario procesado (sin URLs, emojis o duplicados)	Texto libre	Normalizado en minúsculas

Criterios de Inclusión y Exclusión

Tipo de criterio	Condición
Inclusión	Comentarios en español
Inclusión	Referencias explícitas o implícitas a seguridad ciudadana
Inclusión	Publicaciones públicas con ≥ 5 palabras
Exclusión	Comentarios duplicados o idénticos
Exclusión	Publicaciones privadas o restringidas
Exclusión	Contenido automatizado o spam

Procedimiento de Aplicación

Fase	Descripción
Definición de parámetros	Selección de palabras clave, hashtags y páginas de origen
Extracción de datos	Uso de web scraping (Selenium, BeautifulSoup) y API de Twitter/X
Depuración inicial	Eliminación de duplicados, spam y mensajes irrelevantes
Preprocesamiento	Limpieza de texto: normalización y eliminación de caracteres especiales
Clasificación automática	Aplicación de modelos PLN (BERT) para polaridad, emoción y tema
Validación humana	Revisión de una muestra (15%) para control de calidad y coherencia temática
Anonimización	Supresión de identificadores personales y reemplazo por códigos alfanuméricos
Consolidación final	Construcción del dataset definitivo para análisis

Anexo 3. Fragmento del dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Comentario	Sentimiento	Confianza_PO	Confianza_NE	Confianza_NE	Tema	Recomendación	Es_Queja	Es_Urgente	Nivel_Urgencia	Tipo de emoción	Intensidad Emocional	Miedo / Desconfianza	Confianza Institución
2	honestamente, ahora mismo, con este gobierno, la policía es más parecido a los	Negativo	0.01	0.08	0.91	Delito	No	No	Si	Bajo	Indignación	0.8	0.75	0.1
3	se pierde tiempo porque ya tienes una patrulla auxiliando mientras el de puertas	Positivo	0.42	0.53	0.05	Policia	No	No	Si	Bajo	Esperanza	0.7	0.1	0.85
4	de tortura y terrorismo de Estado que ofrece la derecha. No es así como se garantiza la paz y la seguridad ciudadana	Negativo	0	0.02	0.98	Delito	No	Si	No	Bajo	Indignación	0.8	0.75	0.1
5	el metro va volando..... ¿cierto Carlos Nayibe Galán de López?	Negativo	0.01	0.04	0.95	Delito	No	No	No	Bajo	Indignación	0.85	0.85	0.05
6	Carlos Galán Bogotá es urgente	Neutro	0.16	0.62	0.22	Delito	No	No	Si	Alto	Ansiedad	0.7	0.7	0.5
	condcha para mentir encubriendo algo tan importante como es								Si		Indignación	0.9	0.85	0.05