



Pontificia Universidad
JAVERIANA
Cali

Desarrollo de un Indicador de Percepción de Paz mediante Análisis de Sentimiento en Publicaciones Digitales

Laura Natalia Losada Sandoval

ID 8986438

Melanie Michelle Gazabón Mora

ID 8986378

*Proyecto Aplicado para optar al título de
Magíster en Ciencia de Datos*

Director(a)

Abel Álvarez Bustos

Codirector(a)

Carlos Ramírez

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, DICIEMBRE 5 DE 2024

FICHA RESUMEN

Desarrollo de un Indicador de Percepción de Paz mediante Análisis de Sentimiento en Publicaciones Digitales

1. **ÁREA DE TRABAJO:** Investigación social
2. **TIPO DE PROYECTO (Aplicado, Innovación, Investigación):** Aplicado
3. **ESTUDIANTE(S):** Laura Natalia Losada Sandoval, Melanie Michelle Gazabón Mora.
4. **CORREO ELECTRÓNICO:** llosada1@javerianacali.edu.co, mgazabon@javerianacali.edu.co
5. **DIRECCIÓN Y TELÉFONO:** 316 2210490, 316 6985145
6. **DIRECTOR:** Abel Álvarez Bustos
7. **VINCULACIÓN DEL DIRECTOR:** Profesor de la Universidad Javeriana de Cali
8. **CORREO ELECTRÓNICO DEL DIRECTOR:** abel.alvarez@javerianacali.edu.co
9. **CODIRECTOR (Si aplica):** Carlos Ernesto Ramírez Ovalle
10. **GRUPO O EMPRESA QUE LO AVALA (Si aplica):** NA
11. **OTROS GRUPOS O EMPRESAS:** NA
12. **PALABRAS CLAVE (al menos 5):** Indicador de percepción de paz, Análisis de sentimientos, Procesamiento del lenguaje natural, Métodos de extracción de palabras claves, Aprendizaje Supervisado
13. **FECHA DE INICIO:** Febrero del 2024
14. **DURACIÓN ESTIMADA (En meses):** 11
15. **RESUMEN:** Medir la percepción del concepto de paz enmarca una complejidad intrínseca a su naturaleza, ya que se relaciona con aspectos objetivos y subjetivos, aspectos objetivos como la ausencia de conflicto armado, y subjetivos, como la sensación de seguridad. Esta dualidad hace que los métodos convencionales de medición sean insuficientes para capturar la verdadera percepción de la paz en una comunidad. En este contexto, este proyecto titulado “Desarrollo de un Indicador de Percepción de Paz mediante Análisis de Sentimiento en Publicaciones Digitales” se basa en la creciente participación de los periódicos digitales, que reflejan sus opiniones y

perspectivas. Desarrolla un indicador para medir la percepción de paz en Bogotá utilizando un modelo de procesamiento de lenguaje natural y análisis de sentimiento, basado específicamente en BERT. Este enfoque permite cuantificar la percepción de paz mediante la categorización de expresiones relacionadas con este concepto, extraídas de publicaciones en periódicos digitales, capturando matices semánticos y emocionales esenciales para comprender el contexto sociocultural de la ciudad. Aquí es donde radica la importancia del proyecto, puesto que se trata del desarrollo de un método de medición que puede capturar la percepción de paz a partir de datos subjetivos. Como resultado, se obtuvo un listado de temas clave con los que los periódicos digitales relacionan la paz, un modelo de categorización de expresiones y un indicador que cuantifica la percepción de paz derivado de la categorización de expresiones. Este indicador, centrado en la caracterización cuantitativa y medible de la noción de paz, se considera una herramienta útil para la formulación de políticas públicas informadas, permitiendo a los responsables de la toma de decisiones, entender mejor las necesidades y percepciones de la población. Con este proyecto se pretende sentar las bases para el desarrollo de un instrumento aplicable en la evaluación de políticas públicas específicas para promover entornos pacíficos basados no solo en indicadores objetivos, sino también en la percepción real de la población.

TABLA DE CONTENIDO

INTRODUCCIÓN	1
1. DEFINICIÓN DEL PROBLEMA	2
1.1. PLANTEAMIENTO DEL PROBLEMA	2
1.2. FORMULACIÓN DEL PROBLEMA.....	3
2. OBJETIVOS DEL PROYECTO	4
2.1. OBJETIVO GENERAL.....	4
2.2. OBJETIVOS ESPECÍFICOS	4
3. MARCO TEÓRICO Y ANTECEDENTES.....	5
3.1. MARCO TEÓRICO	5
3.1.1. La percepción histórica y contemporánea de la paz	5
3.1.1.1. El concepto de paz.....	5
3.1.1.2. El Índice de Paz Global (Global Peace Index)	7
3.1.1.3. Estudio en Colombia sobre la percepción de paz	9
3.1.2. Metodologías y herramientas utilizadas	11
3.1.2.1. Web scrapping.....	11
3.1.2.2. Minería de datos	12
3.1.2.3. La analítica de texto	12
3.1.2.3.1. Preprocesamiento de datos textuales.....	13
3.1.2.3.2. Procesamiento del lenguaje natural (PNL).....	13
3.1.2.3.3. Similitud de Jaccard.....	14
3.1.2.3.4. Similitud del Coseno para extracción de palabras claves (Similitud de palabras). 14	
3.1.2.3.5. YAKE (Yet Another Keyword Extractor).....	15
3.1.2.4. Análisis de sentimientos con BERT	18
3.1.3. ¿Qué es un indicador?.....	20
3.1.4. ¿Cómo construir un indicador?.....	22
3.2. ANTECEDENTES	26
4. DESARROLLO DEL PROYECTO	31
4.1. OBJETIVO 1	31
4.1.1. Recopilación de la base de datos	31
4.1.2. Análisis descriptivo de la base de datos.....	33

4.1.3.	Preprocesamiento de datos	36
4.1.4.	Identificación de palabras similares	38
4.1.5.	Similitud de Jaccard en el discurso de los periódicos	41
4.1.6.	Análisis de Palabras Clave con YAKE (Yet Another Keyword Extractor).....	43
4.1.7.	Creación del corpus de trabajo	44
4.2	OBJETIVO 2	48
4.2.1	Análisis de sentimiento	48
4.2.1.1	Modelo no supervisado.....	48
4.2.1.2	Modelo Supervisado	49
4.2.1.3	Modelo final	53
4.3	OBJETIVO 3	59
4.3.1	Diseño y Desarrollo de un Indicador	59
4.3.2	Evaluación y Análisis de los Resultados Obtenidos con el Indicador Diseñado	65
4.3.3	Comparación entre el Índice de Percepción de Paz en Bogotá y el Índice de Paz Global	75
5	CONCLUSIONES Y TRABAJOS FUTUROS.....	77
5.1	CONCLUSIONES	77
5.2	TRABAJOS FUTUROS.....	78
6	REFERENCIAS BIBLIOGRÁFICAS	80
7	ANEXO 1	84

LISTA DE ILUSTRACIONES

Ilustración 1 Evolución del índice GPI de Colombia	9
Ilustración 2: Los elementos más importantes de la paz.....	10
Ilustración 3: Definición de "lo contrario de la paz"	10
Ilustración 4 Web scraping.....	11
Ilustración 5 Modelo Bert.....	19
Ilustración 6: Criterios de valoración de indicadores.....	25
Ilustración 7 Distribución del número de palabras en los artículos de paz	34
Ilustración 8 Número de autores por cantidad de artículos de paz	35
Ilustración 9 Número de artículos de paz por mes	35
Ilustración 10: Preprocesamiento de la base de datos.....	38
Ilustración 11: Palabras relacionadas con el término paz.....	39
Ilustración 12 Selección de términos relacionados con la Paz.....	40
Ilustración 13 Selección de términos relacionados con la guerra	41
Ilustración 14:Similitud de Jaccard entre periódicos	42
Ilustración 15 Número de artículos por año de la base de datos	46
Ilustración 16 Distribución del número de palabras en los artículos.....	46
Ilustración 17 Número de autores por cantidad de artículos	47
Ilustración 18: Distribución de clases en base de entrenamiento antes del balanceo	50
Ilustración 19: Distribución de clases en base de entrenamiento después del balanceo	51
Ilustración 20 Proceso final modelo análisis de sentimiento.....	53
Ilustración 21Matriz de confusión Modelo de sentimiento desarrollado	55
Ilustración 22 Distribución de sentimientos por periódico.....	56
Ilustración 23 Análisis de sentimiento total.....	58

Ilustración 24 Comparación de los modelos propuestos como índice de percepción de paz.....	63
Ilustración 25: Evolución del índice de percepción de paz desarrollado	66
Ilustración 26: Registro histórico de homicidios en Bogotá 2010 -2023	73
Ilustración 27: Registro histórico de hurtos en Bogotá 2010 – 2023.....	73
Ilustración 28: Comparación entre el índice de percepción de paz en Bogota y el índice de paz global	75

LISTA DE TABLAS

Tabla 1 Composición de la base de datos con el término <<Paz>>	32
Tabla 2 Cuadro comparativo top 20 palabras similares.....	40
Tabla 3 Composición de la base de datos final	45
Tabla 4: Ficha técnica del indicador desarrollado.....	65
Tabla 5 Línea de tiempo proceso de paz en Colombia 2020 – 2023	71

INTRODUCCIÓN

En el contexto actual, la noción de paz ha evolucionado significativamente, abarcando no solo la ausencia de conflictos bélicos, sino también aspectos complejos que reflejan los valores y aspiraciones de la sociedad. Esta evolución conceptual de la paz se evidencia en los periódicos digitales de Bogotá, que capturan las percepciones, preocupaciones y opiniones de los ciudadanos sobre el tema.

Este proyecto surge ante la necesidad de herramientas efectivas para medir la percepción de paz en Bogotá a través de publicaciones digitales. La tarea se complica debido a la diversidad y matices de las narrativas digitales contemporáneas, lo que limita la capacidad de las autoridades, como la Alcaldía Mayor de Bogotá, la Secretaría Distrital de Seguridad, Convivencia y Justicia, la Secretaría Distrital de Gobierno, y la Defensoría del Pueblo, para comprender profundamente las necesidades y expectativas de la comunidad en el ámbito de la paz. Estas entidades son fundamentales para diseñar políticas públicas que reflejen las verdaderas inquietudes ciudadanas, basadas en un entendimiento preciso de las percepciones sociales.

Para abordar esta carencia, se ha desarrollado un indicador basado en un modelo de análisis de sentimientos y procesamiento de lenguaje natural, basado específicamente en BERT, e identificación de palabras relacionadas con la paz en los periódicos digitales de Bogotá. Este enfoque no solo beneficia a las autoridades locales, sino también a instituciones como la Defensoría del Pueblo, la Policía Metropolitana de Bogotá, y organizaciones internacionales como la Misión de Verificación de las Naciones Unidas en Colombia, ya que permite identificar temas clave sobre la percepción de paz, analizar patrones de positividad, negatividad o neutralidad en relación con estos temas, y cuantificar de manera objetiva y contextualizada la intensidad de estas percepciones.

El proyecto, no solo aspira a contribuir al ámbito académico, sino también a proporcionar a las autoridades mencionadas una herramienta valiosa para la toma de decisiones informadas en políticas públicas. La implementación de este indicador ofrece una comprensión más completa y matizada de la percepción de paz en Bogotá, fomentando un entorno más armonioso y sostenible. Además, el proyecto tiene el potencial de generar un impacto social significativo al fortalecer la participación ciudadana y considerar las voces individuales y colectivas en la evaluación de la paz, contribuyendo así a una sociedad más inclusiva y empoderada, donde las políticas públicas reflejen fielmente las necesidades reales de la población.

1. DEFINICIÓN DEL PROBLEMA

El término “Paz” ha ido transformándose a lo largo de la historia. Los primeros intentos de enfrentarse desde una perspectiva científica a la problemática, datan de los años inmediatamente posteriores a la Primera Guerra Mundial en los Estados Unidos, en donde, se definió el concepto tradicional “Paz” como la mera ausencia de guerra, posteriormente diversas corrientes de pensamiento y movimientos sociales contribuyeron a la evolución del concepto hacia un enfoque más holístico, que se caracteriza por un concepto de “Paz” asociado con otros valores considerados deseables, como la justicia, la libertad y la ausencia de cualquier tipo de conflicto. De modo que lo que se ha entendido por “Paz” varía en función de las coordenadas temporales en las que nos situemos [1].

En la actualidad, se reconoce que el término de “Paz” es polisémico, es decir, que no puede basarse en una definición universal, sino que debe adaptarse a las realidades y valores específicos de cada comunidad. Esta diversidad de interpretaciones refleja la complejidad inherente a su medición. El desafío radica en capturar la riqueza y la subjetividad de las experiencias individuales y colectivas relacionadas con la paz, considerando la diversidad cultural y contextual. La medición efectiva de la paz no solo contribuye a la toma de decisiones informada en políticas públicas, sino que también permite el diseño de estrategias de desarrollo que respeten y fomenten los valores específicos de cada comunidad, promoviendo así un bienestar sostenible y una convivencia armoniosa.

1.1. PLANTEAMIENTO DEL PROBLEMA

Los intentos por cuantificar la paz en la sociedad comenzaron en el siglo XX, cuando los investigadores, partiendo de la definición de paz como ausencia de guerra, utilizaron cifras para caracterizar los conflictos bélicos. Uno de los pioneros en este campo fue Lewis Richardson, quien en las décadas de 1940 y 1950 dedicó sus esfuerzos a recopilar información estadística sobre todos los conflictos documentados desde 1820 hasta mediados de la década de 1940. Su monumental obra permaneció inédita hasta después de su muerte. Fue publicada en 1960 bajo los títulos “*Arms and Insecurity*” y “*Statistics of Deadly Quarrels*”. En esta última obra, Richardson ensaya una clasificación de «todas las luchas mortales» basándose en el número de muertos. Construye de este modo un *continuum* que se extiende desde un muerto (asesinato), hasta diez millones de muertos (Segunda Guerra Mundial). El criterio que permitía hablar de una situación de guerra se situaba en torno a los mil muertos [1].

Actualmente, el índice se utiliza para medir y evaluar diversos aspectos de la sociedad, incluyendo la paz. En este contexto, el Índice Global de la Paz (*GPI*, por sus siglas en inglés) es elaborado y publicado desde 2007 por el *Institute for Economics and Peace*, en colaboración con expertos de institutos para la paz y de la Universidad de Sídney. Los datos son procesados por la Unidad de Inteligencia del semanario británico *The Economist*. Está compuesto por 23 indicadores cuantificables, donde los indicadores de paz interna suponen un 60% del valor y los de paz externa un 40% [2]. Este índice deja de lado la experiencia y percepción subjetiva de la población como

la cohesión social, la confianza interpersonal, la equidad y la sensación de seguridad, elementos que son intrínsecos a la experiencia subjetiva de la paz, lo que no permite capturar completamente la complejidad emocional y social asociada con el concepto de paz. La insuficiencia del enfoque tradicional destaca la necesidad de desarrollar herramientas y métodos que integren la perspectiva subjetiva de las personas. El problema radica en la falta de instrumentos de medición de la paz que refleje aspectos subjetivos que permitan la implantación de políticas públicas informadas y orientadas a promover entornos pacíficos basados en la experiencia y percepción de la población [3].

1.2. FORMULACIÓN DEL PROBLEMA

Las publicaciones sobre la paz en periódicos digitales se volvieron más comunes con el paso del tiempo, y en ellas se pueden identificar las preocupaciones, opiniones y perspectivas de la población sobre diversos temas. Este fenómeno digital creó un vasto registro de la narrativa colectiva, reflejando no solo las noticias y eventos, sino también las emociones y percepciones que subyacían a ellos. En este contexto, surgió la pregunta: **¿Cómo evaluar la percepción de paz en la población de Bogotá, a partir de una categorización de las publicaciones del 2023 en periódicos digitales de acuerdo con su análisis de sentimiento?**

Para desarrollar esta pregunta se plantearon primeramente ¿Cómo se obtenía una base de datos de los temas relacionados con la percepción de paz en Bogotá, basándose en el uso de técnicas de *web scraping* y extracción de palabras clave en los principales periódicos digitales?, ¿Cómo se cuantificaba la percepción de paz a partir de la categorización de expresiones relacionadas con el tema en los periódicos digitales, basado en procesamiento de lenguaje natural y técnicas de análisis de sentimientos? Y finalmente, ¿Cómo se podía construir un indicador que capturara la percepción de paz en Bogotá a partir del análisis de sentimiento? Estas preguntas formaron el núcleo de la investigación destinada a entender y medir la percepción de paz en el entorno digital, donde las respuestas pudieron arrojar luz sobre las complejidades subyacentes en la percepción colectiva de la paz en la era de la información.

2. OBJETIVOS DEL PROYECTO

2.1. OBJETIVO GENERAL

Desarrollar un indicador para medir la percepción de paz en Bogotá, a partir de un modelo basado en procesamiento de lenguaje natural y técnicas de análisis de sentimientos que permita cuantificar la percepción de paz a partir de la categorización de expresiones relacionadas al tema presentes en los periódicos digitales.

2.2. OBJETIVOS ESPECÍFICOS

1. Construir una base de datos con las publicaciones vinculadas con la percepción de paz en los medios digitales de Bogotá entre 2020 y 2023 mediante técnicas de *web scraping* y extracción de palabras claves, estableciendo las bases para el desarrollo de un modelo preciso y representativo para el análisis de sentimientos posterior.
2. Desarrollar un modelo basado en procesamiento de lenguaje natural y técnicas de análisis de sentimiento que permita cuantificar la percepción de paz a partir de la categorización de expresiones relacionadas al tema en las plataformas digitales.
3. Crear un indicador con los resultados del modelo desarrollado, en el objetivo específico número 2, que refleje cuantitativamente la percepción de paz derivada de la categorización de expresiones relacionadas al tema en las publicaciones digitales en Bogotá.

3. MARCO TEÓRICO Y ANTECEDENTES

3.1. MARCO TEÓRICO

La revisión literaria de este trabajo se estructura en tres partes fundamentales que abordan aspectos clave para comprender el enfoque y los resultados de la investigación. En primer lugar, se examina la percepción histórica y contemporánea de la paz. Se analizan los enfoques teóricos y las contribuciones académicas relevantes que han influido en la comprensión de la paz en contextos diversos, para establecer un marco conceptual sólido para la investigación. En segundo lugar, se exploran las metodologías y herramientas utilizadas para la recopilación y análisis de datos en periódicos digitales, destacando su relevancia y aplicabilidad en la investigación. Finalmente, se introduce la importancia de los indicadores como medida para evaluar los aspectos clave de la sociedad. Esta revisión de literatura proporciona el contexto necesario para establecer las bases metodológicas del proyecto de investigación.

3.1.1. La percepción histórica y contemporánea de la paz

3.1.1.1. El concepto de paz

El concepto de paz ha sido el objeto de estudio a lo largo de la historia, su significado ha evolucionado de manera significativa en respuesta a las cambiantes dinámicas sociales, políticas y culturales. El concepto de paz ha sido tanto un ideal aspiracional como un objetivo pragmático, moldeado por las experiencias colectivas y las circunstancias históricas de las poblaciones. En las antiguas civilizaciones el concepto de paz estaba vinculado con el término griego "*eirene*" y el latín "*pax*". En griego, "*eirene*" aludía a la armonía del espíritu y de la mente del ser humano en la constante búsqueda de la perfección. Era una armonía que representaba el orden interior y se entendía como un sinónimo de la palabra "*homonía*", la cual era utilizada para referirse a la ausencia de hostilidades dentro de las ciudades o entre estas. Por otro lado, el término latino "*pax*" hacía referencia a la "relación legal y recíproca entre partidos". Por lo tanto, esta noción se acercaba más al respeto por el orden establecido, con una gran carga social y política que llevaba a describir situaciones de orden y unidad dentro de las ciudades [4].

Posteriormente, en la edad media la iglesia católica modificó la idea latina de *pax*, otorgándole una visión universalista, el término se enmarcó en lo divino y lo teológico, un ideal que se encuentra en el cielo, caracterizado por la búsqueda del camino hacia Dios o la paz espiritual, la regulación de la violencia mediante la proclama de las virtudes teologales, morales, y los pecados capitales, preservando el concepto de paz como ausencia de guerra [5].

Durante la Edad Moderna y el Renacimiento, el capitalismo se expandió significativamente y surgió una revolución en la forma de entender el conocimiento, lo que cambió la percepción del ser humano en el mundo. Este cambio permitió que el concepto de paz evolucionara de una idea principalmente moral y espiritual a una más política y legal. La paz comenzó a asociarse con las relaciones entre los estados, los acuerdos diplomáticos y las leyes internacionales. Thomas Hobbes creía que los estados, al concentrar y limitar la violencia que legítimamente se podía

ejercer, eventualmente frenaban la brutalidad connatural a los seres humanos. En otras palabras, la concentración de la violencia era vista como la única forma efectiva para limitarla. Esto se debía a que el poder coercitivo de los estados intimidaba a sus miembros, y, por tanto, la paz se entendía como el resultado de ese miedo común. Por su parte, John Paul Lederach resumió con claridad el estado de la paz en este momento histórico como una competencia exclusiva de los estados modernos, quienes eran los responsables legítimos de mantener la unidad y el orden internos mediante la fuerza [6].

En el siglo XIX se popularizaron las teorías del idealismo y el liberalismo. Desde esta perspectiva, se planteó que el Estado tiene la responsabilidad primordial de establecer la paz a través de la diplomacia y otros recursos políticos. Además, la guerra debe considerarse como un último recurso y evitarse en lo posible. El argumento fundamental del idealismo-liberalismo radica en la premisa de que los individuos buscan la paz a través de relaciones mutuas y utilizan la diplomacia como herramienta principal para preservarla. Por lo tanto, se argumenta que es crucial la presencia de agentes externos que aseguren la seguridad tanto a nivel nacional como internacional, eliminando así los procesos de anarquía que suelen llevar al conflicto [7]. Este enfoque dio lugar al surgimiento de movimientos pacifistas y organizaciones internacionales dedicadas a la resolución de conflictos, como la Sociedad de las Naciones, precursora de las Naciones Unidas. Estos movimientos y organizaciones jugaron un papel fundamental en la promoción de la paz y la cooperación internacional al influir en la creación de tratados internacionales y conferencias emblemáticas. Por ejemplo, las conferencias de La Haya de 1899 y 1907 sentaron las bases para la regulación de los conflictos armados y el establecimiento de mecanismos de arbitraje, promoviendo así la estabilidad y la resolución pacífica de disputas entre naciones.

El siglo XX estuvo enmarcado por dos guerras mundiales y el surgimiento de organizaciones mundiales que promovían la paz. Un hecho para destacar es la publicación de la obra *"Violence, Peace, and Peace Research"* (Violencia, Paz e Investigación sobre la Paz), publicada en 1969. En ella Galtung desarrolló y explicó estas dos formas de concebir la paz: La «paz negativa» se definiría como simple ausencia de guerra y violencia directa, como la guerra o el conflicto armado, mientras que la «paz positiva» se definiría como ausencia de guerra y violencia directa junto con la presencia de la justicia social, relacionando ésta a la presencia de condiciones sociales, políticas y económicas que promueven la justicia, la equidad y el bienestar de todos los miembros de una sociedad, centrandose su esfuerzo en la construcción de relaciones pacíficas, la resolución no violenta de conflictos y el desarrollo de estructuras que fomenten la cooperación y la inclusión [1].

Durante esta época, surgieron los primeros intentos de cuantificar la paz en la sociedad, motivados por la definición de la paz como ausencia de guerra. Uno de los pioneros en este campo fue Lewis Richardson, quien se dedicó a recopilar información estadística sobre conflictos utilizando fuentes documentales desde 1820 hasta mediados de la década de 1940. Su monumental obra permaneció inédita hasta después de su muerte y fue publicada en 1960 bajo los títulos *"Arms and Insecurity"* y *"Statistics of Deadly Quarrels"*. En esta última, Richardson

desarrolló una clasificación de "todas las luchas mortales" basada en el número de muertos, creando un *continuum* que iba desde un muerto (asesinato) hasta diez millones de muertos (Segunda Guerra Mundial) [1].

El avance de la tecnología de la información y la comunicación ha conectado a las personas globalmente, facilitando el activismo y la movilización ciudadana. Sin embargo, esta interconexión también ha expuesto a la sociedad a amenazas transnacionales como el terrorismo y los conflictos armados. En respuesta a estas dinámicas globales, ha surgido un enfoque más holístico de la paz en el siglo XXI, que reconoce que la paz no se limita solo a la ausencia de guerra, sino que también aborda las causas subyacentes de los conflictos y promueve la justicia social, el respeto a los derechos humanos y el desarrollo sostenible [8]. Adicionalmente, cabe señalar que la paz, abarca múltiples dimensiones y perspectivas. Desde la resolución cotidiana de conflictos y la transformación social de la paz imperfecta hasta la promoción de ideas neoliberales y la hegemonía del mercado en la paz liberal, cada enfoque refleja una faceta única. Sin embargo, emerge una visión crítica en la paz decolonial que desafía las narrativas eurocentristas y anglosajonas, abogando por el reconocimiento de saberes y prácticas de paz marginadas. En este contexto, se prioriza la inclusión de comunidades marginadas como los grupos indígenas, campesinos y palenqueros [9]. Esta noción de paz, que no se limita a la ausencia de conflicto, sino que busca alternativas autónomas de dignidad, resuena en la idea de construir diálogos culturales, sociales y comunitarios donde todas las voces sean valoradas [10].

Desde una perspectiva sociopedagógica, la paz se presenta como una combinación de empatía, no violencia y creatividad, reconociendo la necesidad de optar por la no-violencia en diversas situaciones para lograr una paz integral [11]. Por otra parte, el enfoque de paz territorial resalta la importancia de la participación ciudadana activa y el fortalecimiento de los Estados locales en la construcción de paz [12]. En última instancia, la búsqueda de la paz requiere una ética global que involucre tanto a estructuras institucionales como a individuos, trabajando juntos hacia el bienestar personal y común [13].

3.1.1.2. El Índice de Paz Global (Global Peace Index)

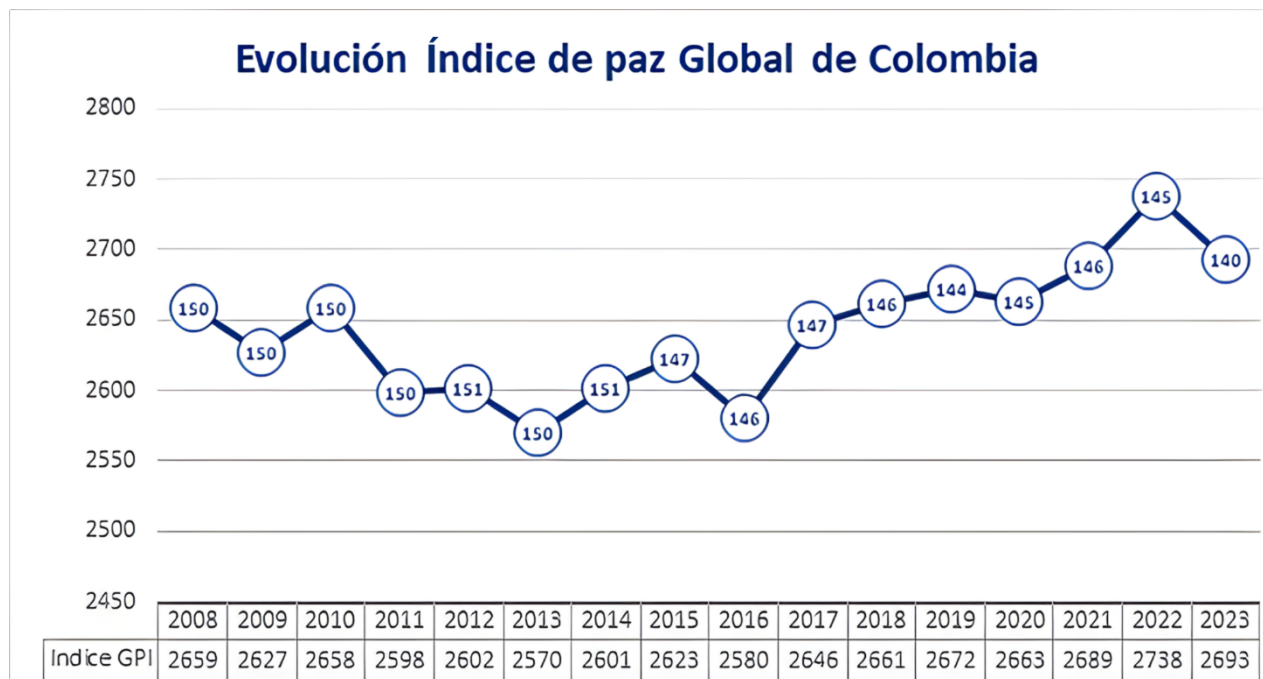
Actualmente, el estudio de la paz es un área fundamental para comprender y abordar los desafíos globales. Una herramienta destacada en este sentido es el Índice de Paz Global (*Global Peace Index*), elaborado y publicado anualmente desde 2007 por el *Institute for Economics and Peace* junto a expertos de institutos para la paz y *think tanks*. Este índice se ha establecido como una referencia clave para evaluar el estado de la paz y la violencia en el mundo. Está compuesto por 23 indicadores cuantificables, donde los indicadores de paz interna representan un 60% del valor, mientras que los de paz externa un 40% [14]. Los indicadores utilizados incluyen:

1. Número de guerras (internas y externas) libradas
2. Muertos en guerras externas estimados
3. Muertos en guerras internas estimados
4. Nivel de conflictos internos
5. Relaciones con estados limítrofes

6. Nivel de criminalidad percibida en la sociedad
7. Número de desplazados por unidad de población
8. Inestabilidad política
9. Nivel de respeto por los derechos humanos (escala de terror político)
10. Posibilidad de actos terroristas
11. Número de homicidios
12. Nivel de criminalidad violenta
13. Probabilidad de manifestaciones violentas
14. Número de personas encarceladas
15. Número de agentes de policía y cuerpos de seguridad
16. Gasto militar con relación al PIB
17. Número de personal militar
18. Importaciones de las principales armas convencionales
19. Exportaciones de las principales armas convencionales
20. Financiamiento de misiones de paz de las Naciones Unidas
21. Número de armas pesadas
22. Disponibilidad de armamento ligero
23. Capacidad o sofisticación militar

A partir de estos indicadores, se evidencia que la paz en una región está correlacionada con varios factores políticos, sociales y económicos, lo que subraya la complejidad del concepto de paz y su relación con diversas dimensiones de la vida en sociedad. Pero hay que considerar que este índice deja de lado la experiencia y percepción subjetiva de la población, aspecto crucial para comprender la realidad de la paz en un contexto determinado. La percepción subjetiva puede proporcionar información valiosa sobre cómo las personas experimentan y perciben la paz en su vida cotidiana, así como sus preocupaciones y temores relacionados con la seguridad y la convivencia. A pesar de esta limitación, el índice sigue siendo una herramienta útil para evaluar y comparar el nivel de paz entre diferentes países, especialmente para observar la evolución de este nivel a lo largo del tiempo y detectar posibles tendencias o cambios significativos en la dinámica de la paz.

Colombia no ocupa una posición favorable en el ranking del Índice de Paz Global (GPI). En 2023, el país estaba en el puesto 140 con 2693 puntos, lo que podría implicar que se considere un lugar de riesgo en términos de seguridad y estabilidad. No obstante, es relevante resaltar que se ha registrado una mejora con respecto al año anterior, cuando se encontraba en el puesto 145 con 2738 puntos [14]. A pesar de este lugar relativamente bajo en el índice, Colombia ha experimentado avances significativos en la mejora de su situación de paz en los últimos años. Es crucial examinar otros factores y tendencias dentro del contexto colombiano para obtener una comprensión más completa de su situación actual en términos de seguridad y estabilidad. Para visualizar la evolución de estos datos en el tiempo, se presenta una gráfica que muestra la posición de Colombia en el ranking dentro de los círculos y los puntos del GPI en el eje Y. Esta representación visual permite observar claramente la tendencia y los cambios en el nivel de paz en el país con los años.



*Ilustración 1 Evolución del índice GPI de Colombia
Elaborada por autores, fuente de información (Datos macro.com, 2022)*

3.1.1.3. Estudio en Colombia sobre la percepción de paz

La empresa Cifras y Conceptos junto con el German *Institute for Global and Area Studies* (GIGA) y la fundación *Friedrich-Ebert-Stiftung* (FESCOL) llevaron a cabo una encuesta nacional durante el 21 de octubre al 11 de noviembre del 2019, con el objetivo de estudiar la concepción de la población colombiana sobre la paz. La muestra incluyó a hombres y mujeres mayores de 18 años de todos los niveles socioeconómicos, residentes en municipios urbanos y rurales en todo el país. Mediante un muestreo probabilístico, multiétnico y estratificado de conglomerados de áreas, con selección aleatoria de la unidad de observación, se obtuvo una muestra de 1733 encuestados, de los cuales 1378 correspondieron a municipios urbanos y 355 a municipios rurales [15]. El cuestionario tenía cuarenta y ocho preguntas diseñadas para explorar la conceptualización de la paz por parte de los encuestados. El bloque principal de preguntas incluía preguntas abiertas y cerradas para captar la comprensión individual de la paz. Además, el estudio abordaba aspectos como la valoración y el conocimiento de diversas instituciones, así como la percepción del marco del Acuerdo de Paz. También se incluyeron preguntas sociodemográficas y sobre ideología política, además de registrar si la persona entrevistada había sido víctima directa del conflicto armado o de la violencia. De esta se pudo determinar:

1. El 51.1% señaló el respeto como uno de los elementos clave de la paz, seguido por tranquilidad (41%), Justicia (39,3%), bienestar (37,2%), derecho (36,2%) y seguridad personal (30,1%) [15].

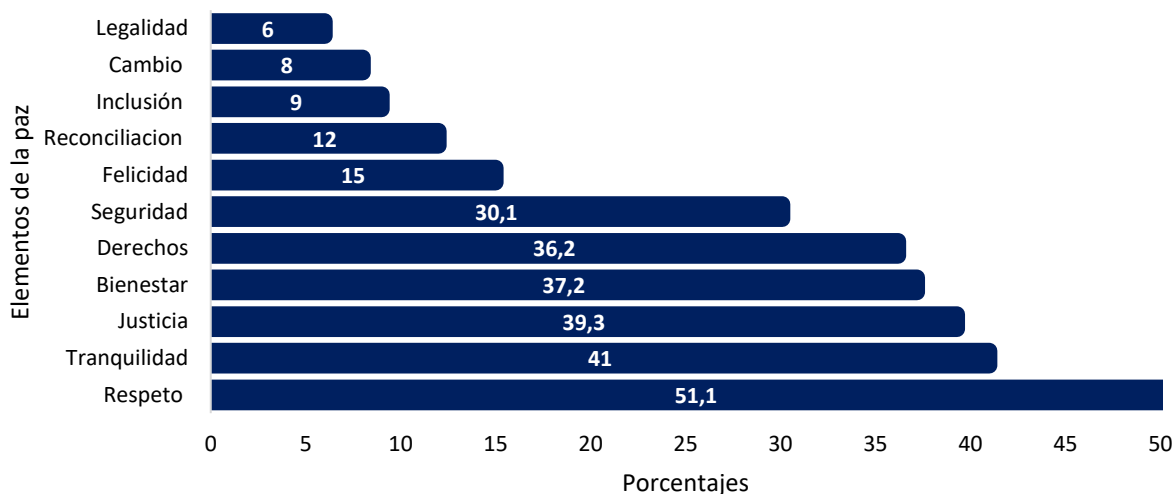


Ilustración 2: Los elementos más importantes de la paz

Fuente: Encuesta de percepción de paz en Colombia basada en la pregunta: <<En su opinión, ¿Cuáles son los tres elementos MÁS importantes de la paz?>> Opciones: Seguridad personal, tranquilidad, respeto, participación, justicia, bienestar, derechos, inclusión, felicidad, legalidad, reconciliación, cambio, educación, salud. Datos ponderados a la estructura sociodemográfica del municipio. Número de observaciones válidas: 1733 [15].

2. El 51,1% relacionó la guerra como lo opuesto a la paz, seguido por la violencia (20,7%). El resto de los atributos como conflicto, intranquilidad, inseguridad e intolerancia fueron mencionados por menos de 10% de la muestra [15].

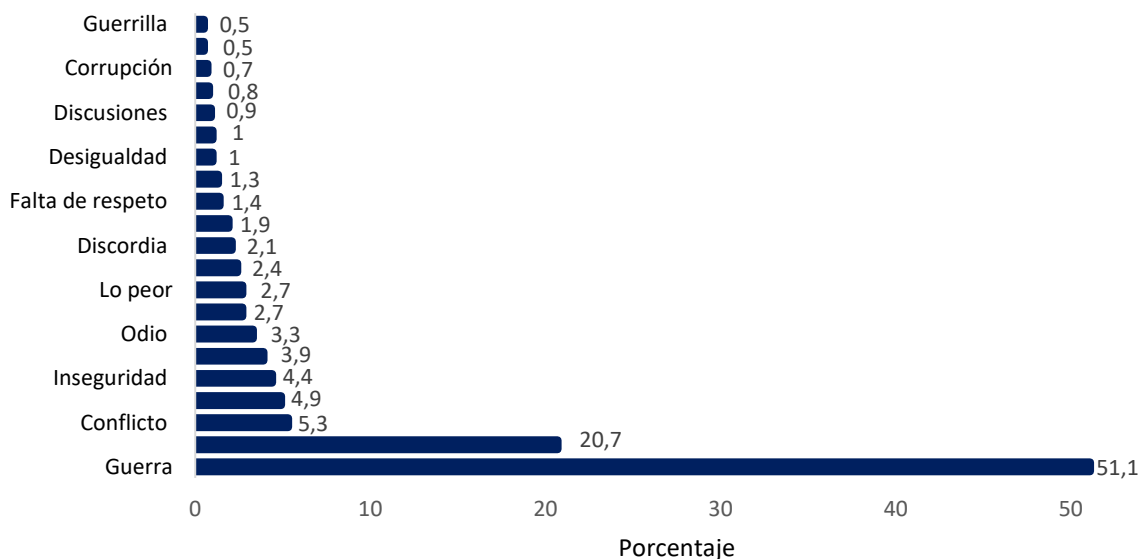


Ilustración 3: Definición de "lo contrario de la paz"

Fuente: Encuesta de percepción de paz en Colombia basada en la pregunta << Complete esta siguiente frase también con lo primero que le venga a la cabeza: Lo contrario de la paz es ...>>. Las respuestas fueron codificadas a posteriori y clasificadas en las siguientes categorías: guerra, violencia, conflicto, intranquilidad, inseguridad, intolerancia, odio, sufrimiento, lo peor, otro, discordia, muerte, indiferencia, guerrilla. Cada respuesta puede ser clasificada en hasta tres categorías. Datos ponderados a la estructura sociodemográfica del municipio. Número de observaciones válidas 1486 [15].

Lo que estos datos nos enseñan es que la mayoría de la ciudadanía en Colombia quiere que cese la violencia y desea gozar de un nivel de tranquilidad en su vida cotidiana. Esto subraya la necesidad de incluir el Análisis de Procesamiento del Lenguaje Natural (PNL) en el estudio de la paz. El PNL permite examinar grandes volúmenes de datos textuales, como opiniones, comentarios y testimonios, para identificar patrones y tendencias en las percepciones y actitudes de la población. Al aplicar el PNL al estudio de la paz, podemos obtener una comprensión más profunda de las preocupaciones y necesidades de la ciudadanía, desarrollando estrategias más efectivas para abordar los conflictos y fomentar la estabilidad en Colombia, adaptadas a las especificidades de las diferentes regiones y comunidades [15].

3.1.2. Metodologías y herramientas utilizadas

3.1.2.1. Web scrapping

El *web scrapping*, también conocido como raspado web, es una técnica ampliamente utilizada para la recopilación de información de sitios web de forma automatizada, utilizando herramientas y técnicas específicas para extraer datos estructurados o no estructurados según sea necesario. Esta técnica se ha convertido en una herramienta poderosa para obtener grandes cantidades de datos de manera eficiente, sin embargo, es importante tener en cuenta las consideraciones éticas y legales al utilizar el *web scrapping*, ya que el acceso y la extracción de datos de sitios web pueden estar regulados por políticas de uso y términos de servicio [16].

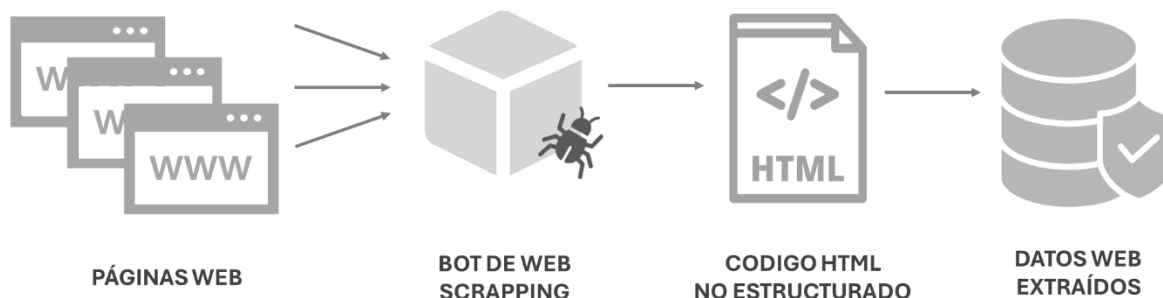


Ilustración 4 Web scrapping

Elaborada por autores basado en el trabajo de grado Desarrollo de un prototipo de aplicación web que permita la extracción de las ofertas laborales de las principales plataformas que postulan empleos en la región caribe, usando la técnica web SCRAPING [10]

Este proceso, tal como se muestra en la ilustración 4, comienza con la identificación de las páginas web de donde se desea extraer la información. Estos sitios pueden contener diversos tipos de datos, como noticias, precios, o información de productos, que a menudo no se presentan de forma accesible para su análisis directo. En la segunda etapa, un bot de web scrapping (programa automatizado) navega por las páginas web de manera similar a como lo haría un usuario humano. El bot solicita información al servidor, como el código HTML de las páginas, y accede a diferentes secciones del sitio según las necesidades del proceso de extracción. Los bots comúnmente se programan usando herramientas como BeautifulSoup, Scrapy, o Selenium, que permiten navegar y obtener datos de manera eficiente [16].

La información extraída inicialmente se encuentra en un formato HTML no estructurado. Este HTML contiene todos los elementos de la página web, pero no está organizado de forma directa para su análisis. En esta fase, el bot debe interpretar y analizar el código HTML para identificar los datos relevantes que se necesitan, como textos, tablas o imágenes. Finalmente, en la última etapa, los datos extraídos se transforman en un formato estructurado, como archivos CSV, bases de datos o JSON, que son más accesibles y fáciles de manejar para su posterior análisis. Este proceso de transformación convierte el código HTML no estructurado en información útil y aprovechable, lo que facilita su integración en proyectos de análisis de datos, reportes, o investigación [16].

3.1.2.2. Minería de datos

La minería de datos emerge como una técnica para extraer información valiosa de grandes volúmenes de datos textuales. Este enfoque permite identificar patrones y relaciones ocultas en los datos que no serían evidentes mediante técnicas más tradicionales de análisis [17]. Al abordar textos extensos y diversos, la minería de datos facilita la estructuración y análisis de información compleja, revelando aspectos cruciales que podrían pasar desapercibidos.

La elección de la minería de datos se justifica por su capacidad para manejar grandes cantidades de datos no estructurados y su habilidad para descubrir patrones significativos a partir de estos. Uno de los principales beneficios de esta técnica es su capacidad para analizar la frecuencia y el contexto de palabras y frases. Mediante técnicas de extracción de información y análisis de patrones, se puede determinar cómo se representan conceptos en diferentes fuentes y cómo varía esta representación entre distintas publicaciones. Este análisis permite capturar no solo la presencia de términos específicos, sino también el contexto y la frecuencia de su uso, lo que proporciona una visión más matizada de la narrativa mediática [17].

Dentro de la minería de datos, el análisis de frecuencia de palabras se destaca como una técnica clave para identificar los términos más comunes en el corpus textual. Este análisis facilita la detección de temas recurrentes y palabras clave [17], ofreciendo una visión clara de las principales narrativas presentes en los textos analizados. Identificar las palabras más frecuentes permite enfocar la investigación en los temas más relevantes. Complementariamente, aunque las nubes de palabras han sido criticadas por ser consideradas decorativas o por su capacidad limitada para ofrecer un análisis profundo de los textos, siguen siendo útiles como una representación visual accesible. Estas nubes destacan la prominencia de términos específicos al mostrar de manera clara los términos más frecuentes y los temas más discutidos. Si bien su valor analítico puede ser debatido, son efectivas para facilitar una primera interpretación de los datos y comunicar hallazgos de manera comprensible a audiencias más amplias, incluyendo aquellas sin formación técnica.

3.1.2.3. La analítica de texto

La analítica de texto es una disciplina que permite enfocarse en la extracción de información útil y la comprensión de patrones a partir de grandes volúmenes de datos textuales [18]. Esto puede

incluir la identificación de palabras clave, la detección de patrones en el uso del lenguaje y la evaluación de la polaridad de los textos. A través de estas técnicas, se logra una comprensión más profunda del contenido y los contextos en los que se presenta [19].

3.1.2.3.1. Preprocesamiento de datos textuales

El preprocesamiento de datos es un paso fundamental que implica la limpieza, transformación y organización de datos brutos para prepararlos adecuadamente antes de aplicar técnicas de análisis [20]. En el contexto del análisis textual, este proceso abarca varias etapas críticas:

1. Limpieza de Datos: Esto incluye la eliminación de palabras comunes (stopwords), caracteres especiales, signos de puntuación, números y otros símbolos que no aportan significado al análisis.
2. Normalización: Consiste en convertir todas las palabras a una forma homogénea, lo que incluye la eliminación de distinciones entre mayúsculas y minúsculas.
3. Corrección Ortográfica: Implica la identificación y corrección de errores ortográficos en el texto.
4. Tokenización: Este proceso divide el texto en unidades significativas, conocidas como tokens, que pueden ser palabras o frases cortas.
5. Lematización: Consiste en reducir las palabras a su forma base o raíz, facilitando el análisis de su significado.

Estas técnicas son esenciales para transformar grandes volúmenes de datos textuales en un formato estructurado y homogéneo. El preprocesamiento asegura que el análisis se realice sobre una base sólida, minimizando la distorsión causada por inconsistencias lingüísticas o elementos irrelevantes. Esto, a su vez, proporciona resultados más precisos y fiables [20].

3.1.2.3.2. Procesamiento del lenguaje natural (PNL)

El procesamiento del lenguaje natural (PLN) es una técnica que permite descomponer y analizar el contenido lingüístico de manera automatizada, proporcionando una visión más estructurada y precisa de cómo se discuten diferentes conceptos en textos. A diferencia de otros métodos de análisis, el PLN se destaca por su capacidad para manejar datos textuales complejos y no estructurados, lo que lo convierte en una herramienta esencial para extraer información significativa de grandes volúmenes de texto [21].

Uno de los principales beneficios del procesamiento de lenguaje natural es su habilidad para realizar análisis semánticos y sintácticos avanzados. El PLN permite la identificación y extracción de entidades, temas y conceptos clave relacionados, así como la clasificación y análisis de las relaciones entre estos elementos [21]. A través de técnicas como el etiquetado de partes del discurso, el análisis de sentimientos y la extracción de temas, el PLN facilita una comprensión más profunda de los temas abordados en el contenido textual. Por ejemplo, puede identificar y

clasificar conceptos clave o detectar patrones que permitan desentrañar las relaciones subyacentes en el texto [21].

El análisis de sentimientos es otra técnica valiosa dentro del PLN, ya que permite cuantificar la polaridad de las menciones sobre determinados temas, como si son percibidas de manera positiva, negativa o neutral. Esta capacidad de medir emociones y tono en grandes volúmenes de texto es crucial para comprender cómo se presentan ciertos conceptos en el discurso público y mediático [21]. El PLN ofrece así una forma eficiente de procesar grandes cantidades de datos textuales, extrayendo información relevante que de otra manera sería difícil de identificar manualmente.

El análisis de similitud de palabras, una técnica utilizada en el procesamiento de lenguaje natural (PLN), permite comprender las relaciones semánticas entre términos dentro de un corpus de datos textuales. Este enfoque facilita la agrupación de términos similares, revelando patrones subyacentes en el lenguaje y destacando cómo diferentes conceptos se interrelacionan [21]. Al descomponer y analizar estas relaciones, es posible identificar conexiones temáticas y estructurales que proporcionan una comprensión más matizada de las narrativas presentes en grandes volúmenes de texto. Esto no solo permite una clasificación más precisa de la información, sino también una visión más profunda de las interacciones entre ideas y conceptos. en grandes volúmenes de texto.

3.1.2.3.3. Similitud de Jaccard

La similitud de Jaccard es utilizada para comparar la similitud entre conjuntos de datos. Se define como el tamaño de la intersección de los conjuntos dividido por el tamaño su unión, expresándose de la siguiente manera [22]:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Donde:

- $|A \cap B|$ representa el número de elementos comunes entre A y B
- $|A \cup B|$ es el número total de elementos en ambos conjuntos.

Esta métrica se basa exclusivamente en la presencia o ausencia de elementos en los conjuntos, sin considerar su frecuencia. Como resultado, el índice de similitud oscila entre 0 y 1, donde un valor de 1 indica la máxima similitud (los conjuntos son idénticos, poseen los mismos elementos) [22].

3.1.2.3.4. Similitud del Coseno para extracción de palabras claves (Similitud de palabras).

La similitud de palabras por coseno se basa en la representación vectorial de palabras, donde cada término es representado como un vector en un espacio multidimensional. Este enfoque permite medir la similitud entre dos palabras calculando el coseno del ángulo entre sus respectivos vectores. Esta medida no solo considera la frecuencia de las palabras, sino también su contexto y

significado, lo que es esencial para capturar matices semánticos que otros métodos de extracción de palabras clave podrían pasar por alto [20].

Una de las razones clave para elegir la similitud de palabras por coseno sobre otros métodos es su capacidad para detectar relaciones semánticas profundas entre palabras. Mientras que los métodos basados únicamente en la frecuencia de palabras pueden identificar los términos más comunes en un texto, no necesariamente capturan la importancia contextual de estas palabras. Por ejemplo, términos menos frecuentes, pero semánticamente relacionados con conceptos clave como "paz" pueden ser igual de importantes para entender la narrativa general. La similitud de palabras por coseno permite identificar estos términos semánticamente relevantes que podrían no ser obvios solo a través de la frecuencia.

La similitud del Coseno mide el coseno del ángulo entre dos vectores positivos no nulos en un espacio con producto interno. Por lo tanto, los vectores de términos que tienen una orientación similar tendrán puntajes más cercanos a 1 ($\cos 0^\circ$), lo que indica que los vectores están muy cerca entre sí en la misma dirección (cerca de un ángulo de cero grados entre ellos). Los vectores de términos con un puntaje de similitud cercano a 0 ($\cos 90^\circ$) indican términos no relacionados con un ángulo casi ortogonal entre ellos. Los vectores de términos con un puntaje de similitud cercano a -1 ($\cos 180^\circ$) indican términos que están completamente opuestos en orientación [20].

Esta técnica es especialmente útil en contextos multilingües o en análisis de textos con variabilidad lingüística significativa. La capacidad de la similitud de palabras por coseno para manejar sinónimos y palabras con significados similares pero diferentes frecuencias de uso son cruciales para asegurar que el análisis sea y preciso [20].

3.1.2.3.5. YAKE (Yet Another Keyword Extractor)

YAKE (Yet Another Keyword Extractor) es un algoritmo de extracción automática de palabras clave propuesto para identificar términos relevantes dentro de un corpus de texto sin requerir entrenamiento previo ni el uso de grandes corpus etiquetados. En lugar de depender de un análisis lingüístico profundo o del uso de grandes recursos léxicos, YAKE se enfoca en la frecuencia de los términos, su distribución en el texto y su contexto, para determinar qué palabras o secuencias de palabras son más relevantes como palabras clave [23]. El algoritmo sigue un proceso estructurado que se puede dividir en cinco pasos clave:

El primer paso en YAKE es el preprocesamiento del texto, un paso esencial para convertir el contenido en un formato adecuado para el análisis automático. Durante esta fase, se lleva a cabo la limpieza del texto, eliminando términos irrelevantes, como stopwords, y convirtiendo el texto en una representación que pueda ser manejada por el algoritmo. Esto incluye procesos como la tokenización y, opcionalmente, la lematización. El preprocesamiento también incluye la segmentación del texto en oraciones y fragmentos (chunks), lo que facilita la identificación de unidades semánticas que son más relevantes como posibles palabras clave. Esta fase establece la base para los pasos posteriores, ya que la identificación de las combinaciones de palabras relevantes depende de la segmentación adecuada y de la eliminación de información irrelevante [23].

Una vez que se han identificado los términos candidatos, el segundo paso implica la extracción de características estadísticas para identificar los términos más relevantes dentro del texto. YAKE se enfoca en las siguientes métricas para caracterizar los términos:

- Frecuencia de término (TF): Es la cantidad de veces que un término o aparece en el texto. Los términos que aparecen con mayor frecuencia suelen ser más relevantes, pero esto no siempre es cierto, ya que pueden ser también términos muy generales como artículos, preposiciones, etc [23].
- Posición: La ubicación de los términos dentro de las oraciones también influye en su relevancia. Los términos que aparecen al inicio o final de las oraciones o que están ubicados cerca de otros términos relevantes, como nombres propios o términos técnicos, pueden ser más importantes [23].
- Frecuencia de co-ocurrencia: Mide qué tan a menudo dos o más términos aparecen en contextos similares dentro del texto. Los términos que tienden a aparecer juntos en el mismo contexto a menudo están relacionados semánticamente y son más probables de ser palabras clave [23].
- Dispersión: La dispersión de un término a lo largo del texto puede ser otro indicador de su relevancia. Los términos que aparecen de manera dispersa en diferentes partes del texto pueden indicar que son temas recurrentes del contenido [23].

Estas características se calculan para cada término identificado en el texto, y se utilizan para determinar cuáles son los términos más relevantes para el contexto.

Una vez extraídas las características estadísticas, el siguiente paso es asignar una puntuación de relevancia a cada término utilizando una combinación heurística de las métricas extraídas. YAKE evalúa la frecuencia de término, la frecuencia inversa del documento (IDF), y la dispersión, para calcular una puntuación global para cada término. Los términos que tienen una alta frecuencia en el texto y una baja frecuencia en el resto del corpus tienen una mayor probabilidad de ser relevantes. La puntuación de cada término refleja su importancia dentro del texto, y los términos con las puntuaciones más altas son seleccionados como las mejores palabras clave. YAKE también realiza ajustes si un término contiene una stopword, ya que las palabras vacías tienden a tener una alta frecuencia, pero poca relevancia semántica [23].

El cuarto paso en YAKE consiste en generar n-gramas, que son combinaciones de términos consecutivos. Mientras que los términos individuales pueden proporcionar información valiosa, las secuencias de términos (n-gramas) pueden capturar significados más complejos y representativos del contenido del texto. Los n-gramas generados pueden ser de cualquier tamaño, desde 1-gramas (palabras individuales) hasta secuencias de mayor longitud (2-gramas, 3-gramas, etc.). Cada n-grama generado se evalúa de manera similar a los términos individuales, tomando en cuenta su frecuencia y co-ocurrencia dentro del contexto del texto. Los n-gramas que contienen combinaciones de términos significativos suelen tener una mayor puntuación de relevancia que las combinaciones de términos más generales [23].

El último paso en el proceso de YAKE es la deduplicación de los términos clave. En este paso, el algoritmo compara términos o n-gramas similares para evitar la redundancia en la lista final de

palabras clave. Utilizando una métrica de similitud de Levenshtein o Jaro-Winkler, YAKE evalúa qué tan similares son dos términos y, si el grado de similitud es alto, se descarta el término redundante. Este proceso permite mantener solo los términos más relevantes y distintivos, eliminando aquellos que no aportan nueva información. Una vez deduplicados, los términos se ordenan en función de sus puntuaciones de relevancia, y se genera una lista final de palabras clave que representan los conceptos más importantes del texto [23].

El desempeño del algoritmo está sujeto a varios factores que pueden influir en la calidad de los resultados. Es importante considerar tanto las fortalezas como las limitaciones de YAKE para entender cómo se adapta a diferentes tipos de textos y contextos. A continuación, se detallan las principales ventajas y desventajas de este algoritmo, lo que permite una evaluación más completa de su aplicabilidad y eficiencia.

Ventajas de YAKE:

- Independencia del lenguaje: YAKE no requiere recursos lingüísticos o corpus etiquetados específicos para funcionar. Esto lo hace aplicable a múltiples idiomas sin necesidad de ajustes complejos, lo que lo convierte en una herramienta universal [23].
- Sin necesidad de entrenamiento previo: A diferencia de otros métodos de extracción de palabras clave que dependen de grandes cantidades de datos etiquetados o de redes neuronales entrenadas, YAKE no requiere entrenamiento previo. Esto lo hace más accesible y menos costoso en términos de tiempo y recursos [23].
- Enfoque basado en características estadísticas: El algoritmo utiliza una combinación de características estadísticas para evaluar la relevancia de los términos. Esto permite identificar palabras clave de manera efectiva sin necesidad de análisis lingüísticos profundos [23].
- Generación de n-gramas: YAKE es capaz de identificar no solo palabras individuales, sino también secuencias de palabras (n-gramas) que representan conceptos clave más complejos, mejorando la precisión en la extracción de frases clave [23].
- Eliminación de redundancias: A través de un proceso de deduplicación basado en la similitud de distancia, YAKE asegura que solo se mantengan las palabras clave más relevantes y distintivas, eliminando términos repetidos y mejorando la calidad de los resultados [23].
- Eficiencia: El algoritmo es rápido y capaz de procesar grandes volúmenes de texto de manera eficiente, lo que es beneficioso en aplicaciones donde se necesita extraer palabras clave de grandes cantidades de datos [23].

Desventajas de YAKE:

- No considera el contexto semántico profundo: YAKE se basa en características estadísticas y de co-ocurrencia, pero no realiza un análisis semántico profundo ni tiene en cuenta las relaciones de significado complejas entre las palabras. Esto puede llevar a la selección de términos que, aunque frecuentes y relevantes a nivel superficial, no sean los más semánticamente significativos en un contexto específico [23].
- Dependencia de la calidad del preprocesamiento: YAKE depende del preprocesamiento

adecuado del texto, como la tokenización, la eliminación de stopwords, y la segmentación de oraciones. Si estas etapas no se ejecutan correctamente, los resultados pueden verse afectados negativamente [23].

- Limitado en textos con muchas ambigüedades: Aunque el algoritmo es eficiente en identificar términos clave en textos bien estructurados, puede ser menos preciso en textos con ambigüedades o con un lenguaje altamente técnico, donde las relaciones entre los términos no son evidentes solo por su frecuencia y co-ocurrencia [23].
- Posible pérdida de información en textos muy complejos: Al centrarse en la frecuencia y la posición de las palabras, YAKE puede pasar por alto términos relevantes en textos con estructuras complejas o con un lenguaje menos directo, como en textos literarios o de alta especialización [23].

En resumen, YAKE (Yet Another Keyword Extractor) es una herramienta eficiente para la extracción automática de palabras clave, que no requiere entrenamiento previo ni recursos lingüísticos específicos. Su proceso basado en características estadísticas permite identificar términos relevantes de manera efectiva, sin la necesidad de un análisis profundo. Aunque tiene algunas limitaciones, como la falta de un análisis semántico detallado, su rapidez y versatilidad la convierten en una opción valiosa para el desarrollo del proyecto.

3.1.2.4. Análisis de sentimientos con BERT

El análisis de sentimientos se emplea como una técnica para detectar y cuantificar las emociones latentes en los textos. Este método clasifica automáticamente el contenido en categorías sentimentales, como positivo, negativo o neutral, ofreciendo una medida cuantitativa de las emociones expresadas en los artículos [24]. Además de los métodos tradicionales de clasificación de sentimientos, como Naive Bayes, Random Forest, Regresión Logística y Máquina de Vectores de Soporte (SVM), la incorporación de modelos de procesamiento del lenguaje natural más avanzados, como BERT (Bidirectional Encoder Representations from Transformers), ha revolucionado este campo. BERT, basado en redes neuronales de transformadores, es capaz de comprender el contexto y la semántica de las palabras en un nivel mucho más profundo que los métodos tradicionales [24].

BERT es un modelo de lenguaje que utiliza una arquitectura de transformers. Se preentrena en grandes cantidades de texto para entender el contexto y las relaciones entre las palabras. Esto se hace bidireccionalmente, considerando tanto las palabras antes como después de una palabra dada. BERT divide el texto en unidades llamadas "tokens" y utiliza múltiples capas de codificadores para procesar la información del texto. Cada capa aplica la atención a diferentes partes del texto. Después del preentrenamiento, BERT se ajusta o "afina" para tareas específicas, como el análisis de sentimientos, alimentándolo con ejemplos etiquetados. Esto le permite adaptarse y mejorar su rendimiento en la tarea específica. En resumen, BERT comprende el contexto y las relaciones entre las palabras en el texto, lo que lo hace efectivo para una variedad de tareas de procesamiento del lenguaje natural, incluido el análisis de sentimientos [25].

Al integrar BERT en el análisis de sentimientos, se puede obtener una comprensión más precisa de las actitudes y emociones expresadas en el texto relacionado con la paz. Esto permite

identificar matices sutiles y capturar el significado completo de las expresiones lingüísticas, lo que resulta fundamental para interpretar de manera precisa el sentimiento en discursos políticos, acuerdos de paz o informes sobre conflictos. El proceso se describe a continuación:

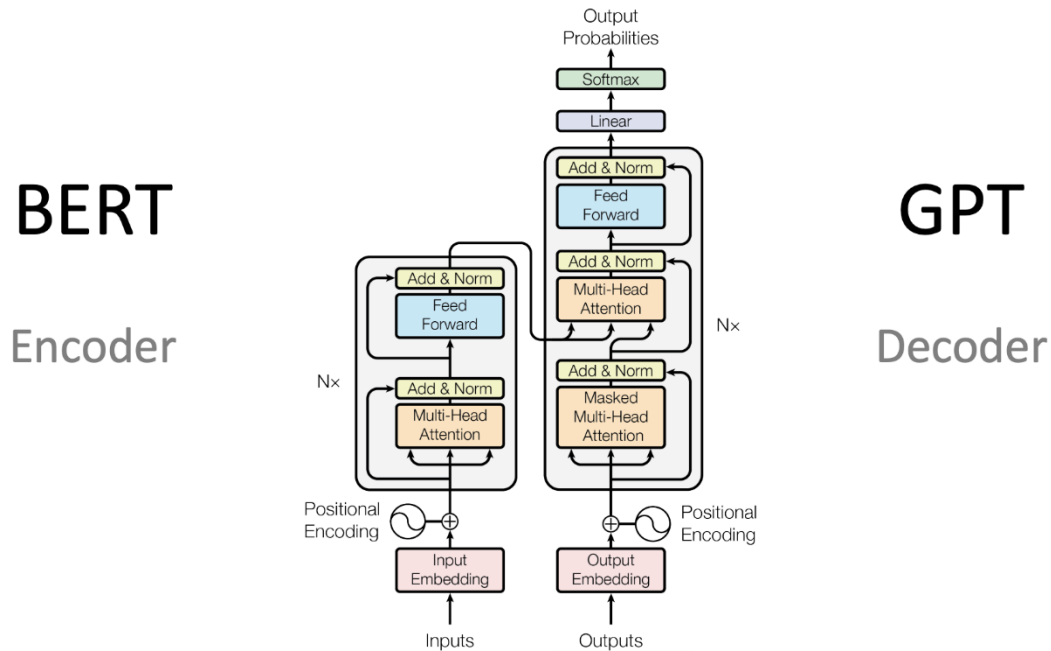


Ilustración 5 Modelo Bert

Fuente: Zhang, Y., Tang, R., & Lin, J.J. (2019). Explicit Pairwise Word Interaction Modeling Improves Pretrained Transformers for English Semantic Similarity Tasks. *ArXiv, abs/1911.02847* [26].

Como se ve la imagen anterior, el modelo BERT recibe un texto y lo tokeniza, dividiéndolo en subpalabras o tokens utilizando el BertTokenizer. Se agregan tokens especiales como [CLS] (inicio de secuencia) y [SEP] (separador). Luego, cada token se convierte en un vector de alta dimensión utilizando embeddings preentrenados, los cuales capturan información contextual sobre cada token, considerando su posición y el contexto dentro de la oración. Estos embeddings pasan a través de varias capas de transformers bidireccionales que capturan relaciones contextuales entre tokens en ambas direcciones (izquierda a derecha y derecha a izquierda), permitiendo al modelo comprender mejor el significado de cada palabra en el contexto de la oración completa.

La primera capa del modelo BERT (en general de cual arquitectura Transformer) tiene como input una inmersión de los tokens en vectores densos para poder pasar de datos textuales no estructurados, a información de tipo numérico que permita alimentar un modelo aprendizaje profundo. Este proceso de inmersión tiene ya implementaciones eficientes y conocidos como Word2Vect, Gloove, entre otras. Sin embargo, BERT utiliza una combinación de tres tipos de embeddings para generar representaciones contextuales de las palabras en una secuencia. Primero, se emplean token embeddings, que son vectores asignados a cada palabra o sub-palabra (según la tokenización de WordPiece), proporcionando una representación básica del significado de cada unidad léxica. A estos se les suman positional embeddings, que indican la posición de

cada token dentro de la secuencia, permitiendo que el modelo entienda el orden de las palabras, ya que los Transformers no tienen un sentido intrínseco del orden. Finalmente, se agregan segment embeddings cuando se trabaja con pares de oraciones, identificando si un token pertenece a la primera o segunda oración. La combinación de estos tres tipos de embeddings (token, posicional y de segmento) forma la entrada que se procesa a través de las capas del Transformer, produciendo embeddings contextuales que capturan el significado de cada palabra en función de su contexto en la secuencia [25].

Finalmente, se toma la representación del token [CLS] (que ahora contiene una representación contextual de toda la oración) y se pasa a través de una capa de clasificación para obtener las probabilidades de cada clase. El modelo produce una distribución de probabilidad sobre las clases definidas (por ejemplo, 'Negativo', 'Neutral', 'Positivo') [26].

La capa de clasificación en BERT para tareas de análisis de sentimientos puede considerarse una regresión logística multiclase. Esta capa final del modelo utiliza una capa densa para generar logits, que son valores continuos no normalizados representando la afinidad de la entrada con cada clase. Estos logits se normalizan mediante la función softmax, que convierte los valores en probabilidades que suman 1, representando la confianza del modelo en cada clase. Esta estructura y proceso son característicos de la regresión logística multiclase, donde la función softmax y la pérdida de entropía cruzada categórica se utilizan para ajustar los pesos del modelo y predecir la clase correcta en problemas de clasificación [26].

El modelo se ajusta (fine-tuning) utilizando un conjunto de datos etiquetado para el análisis de sentimiento. Durante el ajuste, se calcula la pérdida comparando las predicciones del modelo con las etiquetas reales y se utiliza un optimizador para ajustar los pesos del modelo y minimizar la pérdida. El modelo se evalúa en un conjunto de datos de prueba para asegurar su precisión y capacidad de generalización [26].

BERT se destaca frente a otros modelos Transformers debido a su enfoque bidireccional para el entendimiento del lenguaje, lo que le permite capturar con mayor precisión las relaciones contextuales entre las palabras en un texto. En el caso del análisis de sentimientos en español, el modelo BETO, una variante específica de BERT preentrenada con un corpus extenso en español, se presenta como una solución más adecuada. A diferencia de modelos como GPT, que procesan el texto de manera unidireccional, o de versiones generales de BERT que no siempre consideran las peculiaridades lingüísticas del español, BETO ha demostrado un rendimiento superior al manejar construcciones gramaticales, modismos y vocabulario propios de este idioma [26]. Estas ventajas lo convierten en una herramienta más eficaz para identificar matices semánticos y emocionales en textos relacionados con la percepción de paz, permitiendo un análisis contextualizado de entornos hispanohablantes. Esta elección estratégica maximiza la capacidad del modelo para interpretar textos complejos y culturalmente específicos, clave en el desarrollo de métricas precisas y útiles para el análisis [25].

3.1.3. ¿Qué es un indicador?

Es una representación cuantitativa (variable o relación entre variables), verificable objetivamente, a partir de la cual se registra, procesa y presenta la información necesaria para medir el avance o

retroceso en el logro de un determinado objetivo [27]. En la práctica, los indicadores cumplen dos funciones básicas: una de carácter cuantitativo y otra de tipo cualitativo. La primera se centra únicamente en medir el estado de avance de un objetivo, mientras que la segunda cumple más una función valorativa al permitir establecer afirmaciones de carácter calificativo en torno al avance registrado. Ambas funciones son complementarias y permiten responder a preguntas claves relacionadas con la intervención [27].

En cuanto a políticas públicas se refiere, los indicadores se utilizan para realizar seguimiento y evaluación. Estos dos conceptos claves en la gestión pública moderna, si bien son complementarios, tiene significados y alcances diferentes. Por un lado, el seguimiento se define como un proceso continuo en el tiempo que genera información sobre el progreso de una política, programa o proyecto. No examina la causalidad entre las intervenciones y sus efectos. Por su parte, la evaluación es una valoración puntual de una intervención y sus efectos (positivos o negativos, esperados o no) que busca determinar una relación de causalidad entre estos. Esto en términos de eficiencia, eficacia, calidad y economía [27].

Las principales características de un indicador son:

- **Simplificar:** Un indicador es una herramienta de síntesis que permite captar aspectos relevantes de una realidad compleja y multidimensional. Las políticas públicas abarcan una gran variedad de áreas (económicas, sociales, culturales, políticas, entre otras), y el rol de un indicador es seleccionar y resumir las dimensiones clave que se desean evaluar. Aunque no abarca todas las dimensiones posibles, su utilidad radica en facilitar la comprensión y monitoreo de aspectos específicos sin sobrecargar de información al usuario [28].
- **Medir:** Los indicadores cuantifican y permiten la comparación de situaciones específicas en distintos momentos del tiempo o en relación con estándares o patrones predefinidos. Medir facilita evaluar si las políticas están logrando avances, retrocesos o permanecen estables en una dimensión específica [28].
- **Comunicar:** Un indicador debe ser claro y comprensible para transmitir información a los tomadores de decisiones, ciudadanos y otros interesados. Su propósito es guiar la toma de decisiones informadas y apoyar el diseño y evaluación de políticas [28].

Los tipos de indicadores son:

- **Indicador de Gestión:** Este tipo de indicador permite medir y controlar los recursos (físicos, humanos y financieros) invertidos en la implementación de políticas o programas, así como la eficiencia con la que se llevan a cabo las actividades necesarias para lograr los objetivos. Los indicadores de gestión ayudan a evaluar el uso de los recursos y a mejorar los procesos, procedimientos y operaciones involucrados en la etapa de implementación. Son fundamentales para analizar si las acciones se realizan conforme a lo planificado y si se optimizan los recursos disponibles [28].

- **Indicador de producto:** Este indicador mide la cantidad y calidad de los bienes y servicios producidos o entregados como resultado directo de una intervención. Los indicadores de producto se centran en los resultados inmediatos, como la provisión de servicios o productos necesarios para alcanzar efectos directos y específicos en la población objetivo. Existen dos tipos de indicadores de producto: de oferta y de demanda. Los indicadores de oferta son aquellos que miden la capacidad de las entidades para proveer dicho bien y/o servicio, mientras que los indicadores de demanda son los que permiten cuantificar los beneficiarios o receptores de los productos [28].
- **Indicador de resultado:** Son aquellos que cuantifican los efectos relacionados con la intervención pública; dichos efectos pueden ser incididos por factores externos y no necesariamente se producen directamente por la intervención pública. De esta forma, estos indicadores miden los cambios de percepción, conocimiento, condiciones de bienestar, entre otros [28].

3.1.4. ¿Cómo construir un indicador?

La formulación de indicadores no puede realizarse de manera desarticulada ni fuera de contexto. En la práctica, es indispensable que esté enmarcada en al menos un objetivo de intervención pública específica. Esto en cuanto que el indicador por si solo genera información, que de no ser analizada o contrastada con una meta o valor objetivo no sería de mayor utilidad. Asimismo, su incorporación como parte de un programa, política o proyecto específico permite identificar el éxito o fracaso, avance o retraso de dicha intervención pública, facilitando la focalización de acciones correctivas [27].

Para poder construir un indicador, es recomendable que se sigan los siguientes seis pasos:

1. **Revisar la claridad del resumen narrativo:** Los indicadores serán más claros y precisos si el objetivo al que están asociados es claro. Por lo tanto, el primer paso es revisar la claridad del objetivo [37]. Un objetivo es una situación deseada en relación con un problema identificado y que responde a la pregunta ¿Qué se quiere lograr? Estos deben seguir una estructura que se conforma por al menos tres componentes: 1) la acción que se espera realizar; 2) el objeto sobre el que recae la acción; y 3) elementos adicionales descriptivos [28].

Además, para garantizar la claridad y precisión del objetivo, es recomendable validarlo a través de cinco criterios esenciales: 1) **Específico:** que sea claro sobre qué, dónde, cuándo y cómo va a cambiar la situación. 2) **Medible:** que sea posible cuantificar los fines y beneficios. 3) **Realizable:** que sea posible de lograr a partir de la situación inicial. 4) **Realista:** que sea posible obtener el nivel de cambio reflejado en el objetivo. 5) **Limitado en el tiempo:** que se establezca un periodo de tiempo en el que se debe completar cada uno de ellos [28].

Los objetivos deben ser descriptivos y concretos, ya que la claridad es esencial en el diseño

del indicador. Esto determinará los factores clave a considerar en el indicador, que servirán como base para añadir elementos que permitan monitorear el objetivo [29].

2. Identificar los factores relevantes: Los factores relevantes representan los elementos clave del objetivo, describiendo "qué" se quiere lograr y "en quién" se espera dicho logro. Estos factores son el núcleo que define el alcance del objetivo, capturando tanto el logro esperado como el grupo o entidad sobre la que recae ese logro. Para identificar estos factores, es esencial empezar con el objetivo y diferenciar entre el fin (el objetivo último) y los medios (las acciones o métodos para alcanzarlo). Solo el fin contiene tanto el logro esperado como su destinatario, lo cual es fundamental al seleccionar los factores relevantes, ya que los medios no deben considerarse como tales [29].
3. Establecer el objetivo de la medición: Después de identificar los factores relevantes, puede ocurrir que se generen varios indicadores a partir de ellos. En este punto, los responsables deben definir las dimensiones del indicador, atendiendo a los aspectos más importantes que se requiera cuantificar para hacer seguimiento a los logros alcanzados [29].
Aunque no existen criterios absolutos para determinar qué indicadores deben establecerse, pueden considerarse ciertos aspectos generales. El primer aspecto es definir claramente las dimensiones de los indicadores adicionales. Los indicadores no solo permiten a la sociedad conocer los logros alcanzados; también son herramientas clave para el análisis y la toma de decisiones relacionadas con la gestión y los resultados obtenidos. Los indicadores deben proporcionar información relevante sobre los procesos y la efectividad de las acciones implementadas. El segundo aspecto a tener en cuenta es la forma en que se presentan los indicadores y la interpretación de sus valores. Se recomienda mostrar los indicadores de manera positiva, de modo que se refuerce la percepción de los avances logrados. Esto no significa que se deba omitir o manipular información, sino más bien presentar los indicadores como una representación de los resultados alcanzados [29].
4. Plantear el nombre y la formulación de cálculo: Una vez definidos los factores relevantes de la medición y las dimensiones del indicador que se requieren medir en cada uno de los objetivos, es necesario definir un nombre y un método de cálculo para el indicador. Para establecer el nombre y la fórmula de cálculo se debe considerar que: a) El nombre del indicador debe ser claro y relacionarse con el objetivo de la medición. b) El método de cálculo debe ser una expresión matemática definida de manera adecuada y de fácil comprensión, es decir, deben quedar claras cuáles son las variables utilizadas. Presentar los indicadores como una relación entre dos o más variables permite tener el contexto sobre el cual se desarrolló. Los métodos de cálculo más comunes son el porcentaje, la tasa de variación, la razón y el número índice [29].

5. Determinar la frecuencia de medición: No existe un criterio rígido para definir la frecuencia de medición de un indicador; sin embargo, esta se puede ajustar según las necesidades de información. La frecuencia de medición depende de factores como: a) La frecuencia de actualización de los datos de las variables influye en la generación del indicador. b) El requerimiento de información (periodicidad) para monitorear avances o identificar áreas de mejora. Establecer una frecuencia de medición implica un compromiso para actualizar el indicador y sus variables de acuerdo con lo definido, lo cual es fundamental tanto para el monitoreo de objetivos como para la transparencia del programa [29].

6. Seleccionar los medios de verificación: Los medios de verificación son las fuentes donde se encuentra la información necesaria para construir un indicador. Estos medios incluyen documentos oficiales, reportes internos, bases de datos, entre otros. Los medios de verificación permiten evaluar la viabilidad de construir un indicador y asegurar la transparencia. Si el programa genera o consolida la información, entonces se establecen los medios de verificación. Si no, es útil buscar si alguna otra institución produce dicha información. Si la información no está disponible, se debe analizar el costo-beneficio de producirla o considerar un indicador proxy si el costo supera el beneficio. Es fundamental que la frecuencia de medición del indicador coincida con la frecuencia de publicación de los medios de verificación, ya que estos aseguran que las evaluaciones del desempeño del programa sean confiables y accesibles para el público, promoviendo la transparencia [29].

Una vez construidos los indicadores, el siguiente paso es hacer una valoración sobre las características mínimas que deben contener:



*Ilustración 6: Criterios de valoración de indicadores
Elaboración por autores*

1. **Claridad:** Un indicador es claro cuando no presenta ambigüedades en cuanto a lo que se pretende medir. Esto implica que no contiene términos o aspectos técnicos que puedan ser interpretados de múltiples maneras. Es esencial que los términos utilizados sean definidos de manera precisa para asegurar que todos los interesados entiendan el indicador de la misma forma [29].
2. **Relevancia:** Un indicador es relevante si sus elementos principales están directamente relacionados con aspectos fundamentales del objetivo (factores clave). Cuando existe una relación estrecha entre los indicadores y los objetivos, es posible evaluar de manera precisa el logro de los objetivos en aspectos sustantivos [29].
3. **Económico:** La economía de un indicador se evalúa comparando el beneficio de obtener la información necesaria con el costo, tanto económico como humano, de generarlo. Un indicador es económico cuando el beneficio supera el costo asociado. Además, es importante

que el costo de generar el indicador se considere en relación con el presupuesto anual disponible. Es importante notar que un indicador que no es relevante ni adecuado no debería considerarse económico, sin importar su costo de obtención [29].

4. **Monitoreable:** Un indicador es monitoreable si cuenta con medios de verificación claros y un método de cálculo replicable, permitiendo comprobaciones independientes. Para garantizar la monitoreabilidad, el método de cálculo debe ser claro, y la información necesaria para construir el indicador debe estar bien definida en los medios de verificación. Estos medios deben ser lo suficientemente precisos para evitar dudas sobre dónde se puede encontrar la información requerida para construir o replicar el indicador [29].
5. **Adecuado:** Un indicador es adecuado cuando proporciona una base suficiente para evaluar el desempeño de lo que se está midiendo. Esto implica que la información que ofrece es relevante y apropiada para describir los logros alcanzados en un periodo determinado, permitiendo tomar decisiones informadas sobre el progreso y los resultados obtenidos [29].
6. **Aporte marginal:** Este criterio se aplica cuando un objetivo tiene dos o más indicadores. Un indicador tiene aporte marginal si, además de cumplir con los criterios anteriores, proporciona información adicional que no es cubierta por otros indicadores. Esto significa que aporta una nueva perspectiva o medición objetiva que enriquece el monitoreo del desempeño, permitiendo una evaluación más completa y detallada de los resultados [29].

Cabe la posibilidad de se establezcan indicadores que cumplan con las características de claridad, relevancia, economía, adecuación, monitoreabilidad y aporte marginal, pero que sus objetivos definidos no reflejen los resultados que pretende alcanzar. En este sentido, y como se ha señalado de modo reiterativo, para que los indicadores sean los apropiados para el monitoreo y la evaluación de un programa, es necesario que sus objetivos reflejen con precisión los resultados que pretende lograr en el corto y el mediano plazo [29].

El monitoreo y la evaluación de los resultados de cualquier iniciativa o proceso requieren herramientas que faciliten la medición precisa de los objetivos establecidos. Los indicadores desempeñan un papel fundamental en este contexto, ya que permiten sintetizar la información relevante de situaciones complejas, facilitando la toma de decisiones y el seguimiento de los avances. Sin embargo, la efectividad de los indicadores depende de su alineación con los objetivos a evaluar, de manera que proporcionen información relevante, comprensible y útil [29].

3.2. ANTECEDENTES

3.2.1. Tesis de maestría “Análisis de sentimientos en Twitter” desarrollado por José Carlos Sobrino Sande [30]

La tesis presenta los fundamentos teóricos sobre los que se asienta el análisis de sentimientos, su historia, aplicaciones y su relación con el procesamiento del lenguaje natural. En ella se demuestra que las herramientas de clasificación de sentimientos se pueden usar para saber qué es lo que opinan otras personas sobre prácticamente cualquier tema. La cuestión clave es saber qué mensajes son los que se deben capturar, someterlos al proceso de clasificación y analizar la tendencia del sentimiento de manera acertada [30]. El caso de estudio presentado corresponde a la implementación de un clasificador de sentimientos para los mensajes de Twitter basado en algoritmos de aprendizaje supervisados, el cual obtuvo una precisión del 71.79% mediante máquinas de soporte vectorial. El modelo implementado usó un corpus proporcionado por el “Taller de Análisis de Sentimiento (TASS)” de la “Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)” y clasificó los mensajes en 4 categorías: Positivo, Neutro, Negativo y Sin sentimiento. Mediante el siguiente proceso: 1. Corpus 2. preprocesamiento, 3. tokenización, 4. Extracción de las características, 5. reducción de las características, 6. Ponderación de las características, 6. Entrenamiento, 7. Clasificación.

3.2.2. Proyecto “Modelo de análisis de sentimientos en redes sociales con un modelo de minería de datos y machine learning ICETEX” [31]

Este trabajo se centra en realizar un análisis de sentimientos sobre el ICETEX, una entidad estatal colombiana que promueve la educación superior a través de créditos educativos. La percepción negativa de los ciudadanos puede afectar la toma de decisiones sobre el acceso a los servicios del ICETEX, por lo que se propone utilizar el procesamiento de lenguaje natural (NLP), la lingüística computacional y el análisis de texto para extraer información de redes sociales y medios de comunicación [31].

Se utilizó un modelo de clasificación conocido como *Random Forest* para analizar los sentimientos en los tweets relacionados con "ICETEX". Este modelo, entrenado previamente, emplea un conjunto de árboles de decisión para clasificar automáticamente los tweets en categorías de sentimiento, como positivo, negativo o neutro.

Al aplicar este modelo a los tweets recopilados, se obtuvieron resultados que muestran la probabilidad de que cada tweet pertenezca a cada una de las categorías de sentimiento. Además, se asignó una clasificación final de sentimiento a cada tweet, lo que permite identificar patrones y tendencias en los sentimientos expresados en la discusión sobre "ICETEX" en Twitter.

3.2.3. Análisis de sentimientos sobre la percepción de los participantes de una feria virtual de emprendimiento durante el confinamiento [32]

El objetivo de este trabajo fue desarrollar un estudio de análisis de sentimiento y texto para evaluar la percepción de los asistentes a una feria virtual de emprendimiento. Esta feria fue diseñada y adaptada para la modalidad virtual con el respaldo de las Tecnologías de la Información y Comunicación (TIC), como parte del curso de “Emprendimiento Empresarial” del Programa de Ingeniería Química de la Universidad de Cartagena [32]. El desarrollo de este trabajo se llevó a cabo en cuatro fases metodológicas: diseño y organización de la feria, preparación y

adaptación de los datos de las encuestas de percepción, análisis de texto y, finalmente, análisis de sentimientos sobre los resultados de las preguntas cualitativas.

Este estudio se llevó a cabo utilizando herramientas de procesamiento de lenguaje natural (NLP) y visualización de datos para analizar las opiniones de los evaluadores sobre el formato propuesto para una feria virtual. A partir de las opiniones recopiladas, se generó una nube de palabras compuesta por 401 términos relevantes, excluyendo artículos, pronombres y preposiciones. Este análisis se realizó utilizando las bibliotecas nltk para procesamiento de lenguaje natural, Wordcloud para generación de nubes de palabras y matplotlib para visualización de datos en Python.

La nube de palabras destaca los términos más frecuentes en las opiniones de los evaluadores, resaltando palabras como "excelente", "idea", "interesante", "estudiantes", "emprendimiento" y "formato". Estos resultados sugieren una recepción positiva por parte de los evaluadores hacia el formato de la feria, enfatizando su carácter innovador y creativo. En resumen, el estudio revela una percepción favorable hacia el formato propuesto para la feria virtual, respaldado por las opiniones expresadas por los evaluadores a través del análisis de la nube de palabras.

3.2.4. Offensive Language Detection in Turkish Tweets with BERT Models [33]

Este trabajo tiene como objetivo de mejorar la detección de lenguaje ofensivo en tweets en idioma turco utilizando modelos BERT (Bidirectional Encoder Representations from Transformers). Los autores adoptaron una metodología basada en el uso de aprendizaje supervisado con pre-entrenamiento de modelos BERT, que previamente se habían probado con éxito en otras lenguas y tareas de procesamiento de lenguaje natural (PLN). La recolección de datos consistió en la creación de un conjunto de tweets etiquetados manualmente como ofensivos o no ofensivos, los cuales fueron divididos en datos de entrenamiento y prueba para alimentar el modelo. Los modelos fueron entrenados para clasificar los tweets, y los resultados de las predicciones se compararon con un conjunto de datos ya etiquetado para medir el desempeño.

Los resultados del estudio muestran que el modelo BERT superó significativamente a otros métodos convencionales en la detección de lenguaje ofensivo, alcanzando una precisión de más del 90% en la clasificación de los tweets en idioma turco. Entre los hallazgos más destacados se incluye el hecho de que BERT es capaz de manejar con mayor eficacia las sutilezas y complejidades del idioma, lo que lo convierte en una herramienta poderosa para el procesamiento del lenguaje en diferentes lenguas. La investigación también reveló que los modelos entrenados con datos multilingües mostraron una mejora moderada en comparación con los modelos monolingües, destacando la importancia de los datos en múltiples lenguas para mejorar la generalización [33].

3.2.5. Complexity Analysis of Sustainable Peace: Mathematical Models and Data Science Measurements [34]

El objetivo de esta investigación fue analizar los factores que sostienen la paz a través de modelos matemáticos y técnicas de ciencia de datos, utilizando un enfoque inspirado en los sistemas

complejos. Se desarrollan diagramas de bucles causales (CLD, por sus siglas en inglés) y modelos de ecuaciones diferenciales para evaluar las dinámicas de la paz sostenible y los atractores del sistema. A partir de estos modelos, se estudia cómo los factores de la paz interactúan y cuáles son los más influyentes para predecir el estado del sistema a largo plazo. Además, se utilizan métodos de ciencia de datos, como análisis de redes y datos sociales de Twitter, para medir valores cuantitativos de los factores de paz.

El estudio emplea técnicas de minería de datos y análisis de sentimiento en redes sociales, específicamente Twitter, para cuantificar factores relacionados con la paz, como la reciprocidad intergrupala positiva y negativa (PIR y NIR) y las expectativas o memorias históricas de los conflictos. Estos análisis son fundamentales para medir interacciones entre grupos sociales y extraer datos que ayuden a entender mejor las dinámicas de paz y conflicto en sociedades diversas [34]. Para llevar a cabo el análisis de Twitter, el equipo de investigación seleccionó hashtags relacionados con grupos que representaban posturas opuestas, por ejemplo, *#refugeeswelcome* (a favor de los refugiados) y *#BritainFirst* (en contra de los refugiados) en Londres. Mediante scripts de Python y el uso de la API de Tweepy, se recopilaban tweets geolocalizados de usuarios que interactuaban con estos hashtags, y luego se analizaron sus contenidos para medir la interacción intergrupala en contextos sociales más amplios, como eventos deportivos. Los resultados mostraron que los usuarios de ambos grupos interactuaban de manera positiva en algunos contextos, como el apoyo a equipos de fútbol locales, lo cual fue interpretado como un indicador de reciprocidad intergrupala positiva (PIR), a pesar de las diferencias ideológicas. Este hallazgo fue verificado mediante pruebas estadísticas que mostraron que no había diferencias significativas en el apoyo a los equipos deportivos entre los usuarios de ambos grupos.

Además, los investigadores llevaron a cabo un análisis de sentimiento de los tweets, clasificando los mensajes según si contenían palabras relacionadas con el pasado o el futuro. Utilizando la librería nltk de Python para analizar el tono emocional de las publicaciones, se midió el sentimiento positivo y negativo de los tweets. Los resultados indicaron que los tweets que hacían referencia al futuro tenían un sentimiento más positivo que los que hacían referencia al pasado, lo que refleja una percepción más optimista del futuro en comparación con la memoria histórica de los conflictos. Esto es consistente con la hipótesis del estudio, que sugiere que las expectativas juegan un papel crucial en la promoción de interacciones positivas y la construcción de la paz [34]. Los resultados de esta investigación ofrecen un marco metodológico valioso para medir la paz desde una perspectiva cuantitativa y dinámica.

3.2.6. Midiendo la Paz: Iniciativas, Limitaciones y Propuestas [35]

El documento *Midiendo la Paz: Iniciativas, Limitaciones y Propuestas*, publicado por el Institut Català Internacional per la Pau (ICIP), recoge las discusiones y propuestas generadas durante un seminario sobre la medición de la paz, celebrado en Barcelona en 2010. En este seminario, investigadores y especialistas abordaron los desafíos metodológicos, éticos y prácticos en la construcción de indicadores de paz. Entre los temas clave, se discutieron las dificultades para

capturar la paz de forma objetiva debido a su naturaleza multidimensional y la inevitable carga de subjetividad en su definición y medición. Los participantes propusieron una estructura para los indicadores que incluyera tanto factores cuantitativos como cualitativos, abarcando aspectos de paz "positiva" (como el bienestar y la justicia social) y paz "negativa" (la ausencia de conflicto directo). Además, se enfatizó la importancia de medir la paz interna y la contribución de los países a la paz global, utilizando enfoques que van desde la seguridad económica y personal hasta la sostenibilidad ambiental y los derechos humanos [35].

Este marco y sus hallazgos son de gran valor para el desarrollo de un indicador de percepción de paz en Bogotá, ya que subraya cómo los enfoques tradicionales para medir la paz pueden beneficiarse de una perspectiva adaptada a contextos específicos. En el caso de Bogotá, este indicador se centraría en la percepción ciudadana sobre factores que afectan su sensación de paz. La subjetividad inherente a la percepción de paz, discutida ampliamente en el documento, resulta crucial para este proyecto, ya que permite considerar las diversas experiencias y valoraciones de los ciudadanos, elementos que no siempre quedan reflejados en estadísticas convencionales.

4. DESARROLLO DEL PROYECTO

La presente investigación se enfocó comprender y cuantificar la percepción de paz en la población de Bogotá, mediante el análisis de expresiones encontradas en los periódicos digitales durante el 2023. Para abordar esta tarea, se recopilieron los datos a través de técnicas de web scraping y extracción de palabras clave. Después, se implementó un modelo de análisis de sentimiento diseñado para clasificar expresiones relacionadas con el tema de la paz, permitiendo así generar un indicador cuantitativo que refleje la percepción de la población. Este enfoque no solo busca proporcionar una visión amplia y detallada de la perspectiva de paz en Bogotá, sino que también aspira a ofrecer una herramienta útil para la toma de decisiones en el ámbito de las políticas públicas, al proporcionar información precisa y detallada sobre las complejidades emocionales asociadas con el concepto de paz en la sociedad bogotana.

4.1. OBJETIVO 1

Construir una base de datos con las publicaciones vinculadas con la percepción de paz en los periódicos digitales de Bogotá entre 2020 y 2023 mediante técnicas de web scraping y extracción de palabras claves, estableciendo las bases para el desarrollo de un modelo preciso y representativo para el análisis de sentimientos posterior.

4.1.1. Recopilación de la base de datos

El conjunto de datos se construyó a partir de información extraída de los cinco periódicos digitales más relevantes de la ciudad de Bogotá, según el ranking proporcionado por SCImago Lab [36]. Este ranking abarca prensa impresa y digital. Los criterios de selección priorizan la presencia y autonomía de los cybermedios con dominio propio, excluyendo subdominios y aquellos que carecen de autonomía. Es importante destacar que todo el análisis se enfoca en el contenido generado por opinadores, periodistas y editoriales de estos medios, lo cual implica que las percepciones analizadas corresponden únicamente a las expresadas en estas publicaciones. No se trata de un muestreo poblacional dirigido a capturar directamente la opinión de la ciudadanía, sino que las interpretaciones son extrapoladas a partir de los discursos presentes en los medios seleccionados. La clasificación se centra en los medios con proyección nacional-global, regional y 'global', dándole prioridad a aquellos de carácter generalista y universal en su oferta de contenido. Esto significa que se consideran principalmente los medios de comunicación que tienen una amplia audiencia tanto a nivel local como internacional y que ofrecen información variada y relevante para una audiencia diversa.

Durante el proceso de recopilación de datos, se visitó cada página web de los periódicos del ranking de SCImago Lab del 2023 correspondiente a la ciudad de Bogotá. En el buscador interno de cada sitio, se realizó una consulta utilizando el término "paz" y los filtros correspondientes para obtener únicamente las columnas de opinión del 2023. Posteriormente, se realizó una verificación manual de los archivos obtenidos, extrayendo los enlaces de interés. Se aseguró que cada archivo estuviera relacionado con el tema de paz, considerando no solo la mera aparición

del término. A continuación, se empleó la técnica de web scraping o raspado web para recopilar la información de los sitios web de forma automática.

Los periódicos seleccionados para este estudio son, **El Espectador**, **Portafolio**, **La República**, **Semana y Las 2 Orillas**. Todos estos son medios de comunicación de gran relevancia en el contexto colombiano, especialmente en la ciudad de Bogotá. Cada uno de estos periódicos posee un enfoque particular en su cobertura, lo que permite obtener una diversidad de opiniones y perspectivas sobre el tema de la paz. **El Espectador**, conocido por su enfoque investigativo y crítico, proporciona un análisis profundo y reflexivo. **Portafolio**, centrado en temas económicos, aporta una visión de la paz desde la perspectiva de la estabilidad y el desarrollo económico. **La República**, otro referente en el ámbito económico, complementa esta visión con su propia interpretación. **Semana**, con su enfoque en el análisis político y social, ofrece un contexto amplio y variado sobre la paz. Finalmente, **Las 2 Orillas**, un medio alternativo y de opinión, brinda una plataforma para voces menos convencionales y más diversas. La decisión de centrarse en estos cinco periódicos se basó en su relevancia y la calidad de la información relacionada con la paz que proporcionaban dentro del contexto del estudio.

En este estudio, se desarrollaron y ejecutaron scripts personalizados que visitaron las páginas web de los cinco periódicos seleccionados. A través de estas herramientas, se extrajeron automáticamente 94 artículos seleccionados previamente. La información extraída incluyó el nombre del periódico, el título del artículo, el autor, la fecha de publicación, el número de palabras y el texto completo del artículo. La combinación de estos periódicos permite una visión holística y representativa del discurso sobre la paz en Bogotá durante el año 2023. A continuación, se presenta una tabla con la conformación de la base de datos.

	NOMBRE DEL PERIÓDICO	NÚMEROS DE ARTÍCULOS
1	EL ESPECTADOR	31
2	PORTAFOLIO	15
3	LA REPÚBLICA	7
4	SEMANA	19
5	LAS 2 ORILLAS	22
	TOTAL	94

*Tabla 1 Composición de la base de datos con el término <<Paz>>
Elaborada por autores*

El proceso de extracción de datos comenzó con la importación de las librerías necesarias. Requests es una biblioteca de Python que permite realizar solicitudes HTTP de manera sencilla y eficiente, facilitando la interacción con servidores web para obtener contenido y enviar datos mediante métodos como GET y POST. Por otro lado, BeautifulSoup facilita el análisis y la extracción de datos de páginas HTML y XML, permitiendo navegar por la estructura del documento y buscar elementos de interés.

Se utilizó también el módulo CSV, estándar en Python, para trabajar con archivos en formato CSV (Comma-Separated Values), el cual es un formato de archivo de texto plano donde los datos están separados por comas u otros delimitadores. Además, se empleó el módulo OS para interactuar con el sistema operativo, lo que permite trabajar con rutas de archivos y directorios, así como crear y modificar archivos.

El proceso continuó definiendo la ruta del archivo CSV donde se almacenarían los datos extraídos, utilizando `os.path.join()`. Se estableció que la ubicación del archivo se encontraría en el escritorio del usuario. A continuación, se creó una lista llamada `BASE_DATOS`, que contiene las URL de las páginas web de las cuales se extraerían los datos, cada una correspondiente a un artículo o entrada de opinión.

El código itera sobre cada URL en la lista `BASE_DATOS`, realizando una solicitud HTTP GET para obtener el contenido HTML de la página. Luego, mediante BeautifulSoup, se analiza el HTML obtenido para extraer el texto del artículo, buscando un elemento HTML específico que contenga el texto, el cual puede variar según la estructura de la página web. Posteriormente, se unen todos los textos del artículo en una sola cadena y se añaden a la lista `datos_extraidos`.

Una vez extraídos todos los datos, el siguiente paso fue escribirlos en un archivo CSV. Se utilizó `open ()` para abrir el archivo en modo de escritura y `CSV.writer` para registrar cada artículo en una fila del archivo. Este proceso se repitió para cada fuente de noticias seleccionada, incluyendo El Espectador, Portafolio, La República, Semana y Las 2 Orillas.

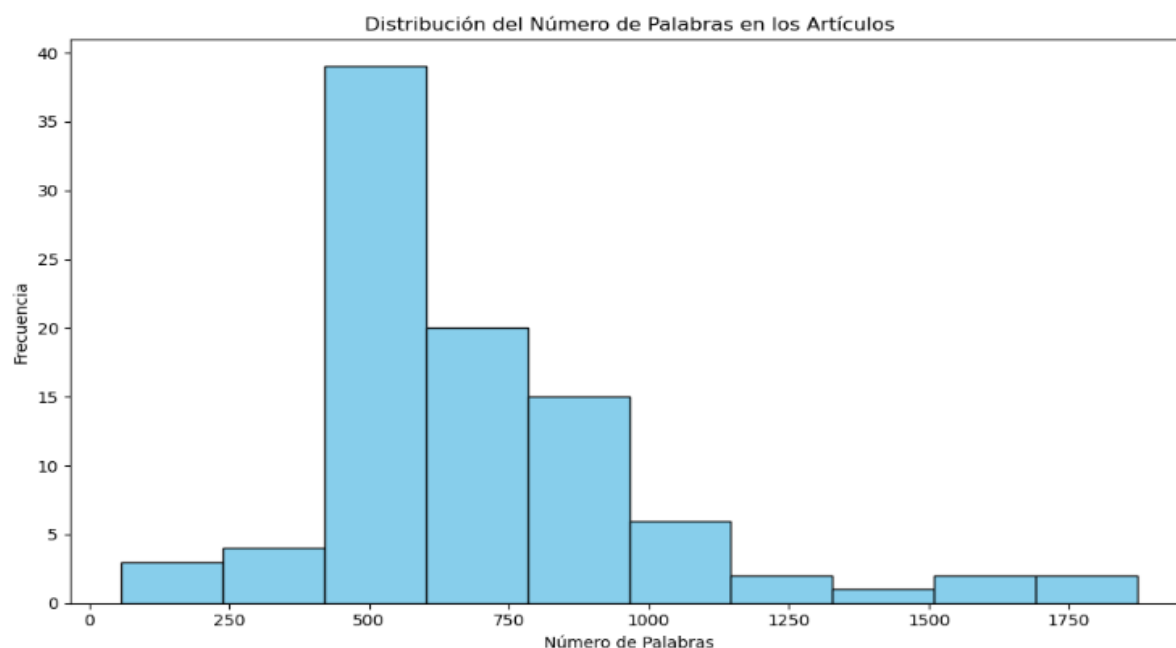
Finalmente, se consolidaron los datos extraídos de cada uno de los archivos CSV en uno solo. Utilizando una lista predefinida con los nombres de los archivos, se abrió cada uno, se leyeron sus datos y se agregaron a una lista general. Estos datos consolidados se escribieron en un nuevo archivo CSV, unificando así la información dispersa en múltiples archivos y facilitando su manejo y análisis posterior.

La utilización de las columnas de opiniones desempeña un papel fundamental en este trabajo, ya que permite identificar y comprender las diversas perspectivas y opiniones de los usuarios de los principales periódicos de la ciudad de Bogotá en relación con el tema de la paz. Estas columnas representan un espacio donde los ciudadanos expresan libremente sus puntos de vista, reflexiones y preocupaciones, brindando una ventana única hacia el sentir colectivo de la sociedad. Al analizar estas opiniones, se puede obtener una visión más completa y contextualizada de la percepción de paz en la ciudad, lo que resulta crucial para comprender las dinámicas sociales, políticas y culturales que influyen en la construcción del concepto paz. Además, este enfoque permite capturar la diversidad de voces y experiencias dentro de la comunidad, promoviendo así un análisis inclusivo y representativo que enriquece la investigación.

4.1.2. Análisis descriptivo de la base de datos

A continuación, se presenta la ilustración 6 que muestra la distribución del número de palabras en los artículos de opinión analizados. Esta ilustración permite visualizar cómo se distribuyen las

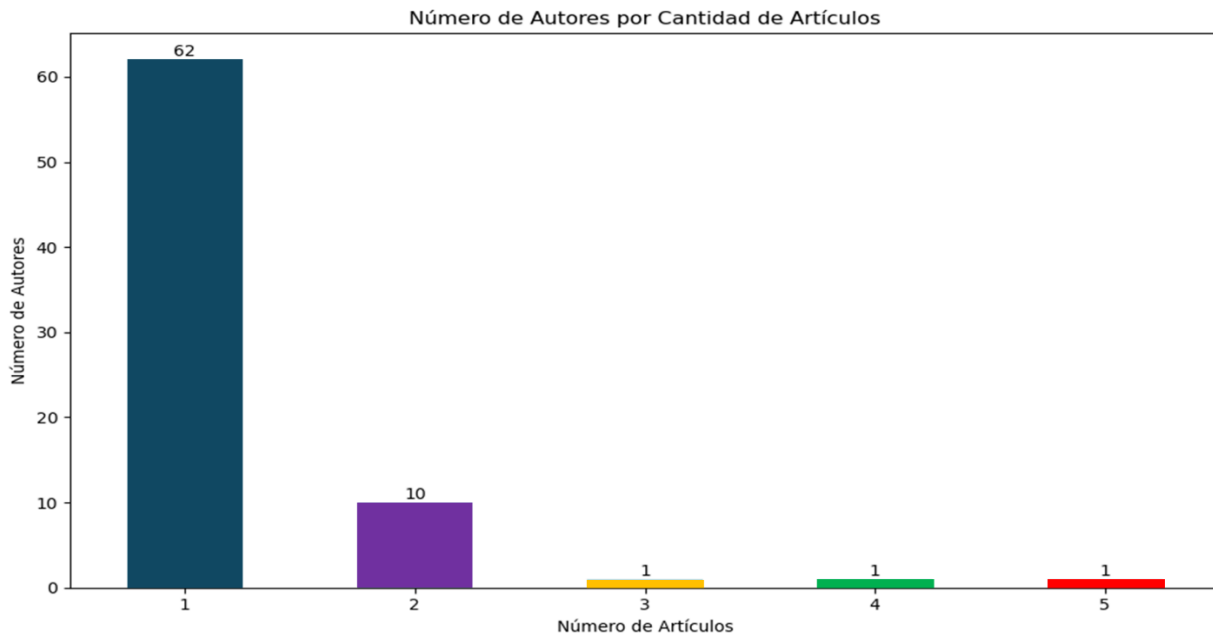
longitudes de los artículos, proporcionando una comprensión más clara sobre la extensión de los textos que abordan el tema de la paz



*Ilustración 7 Distribución del número de palabras en los artículos de paz
Elaboración por autores*

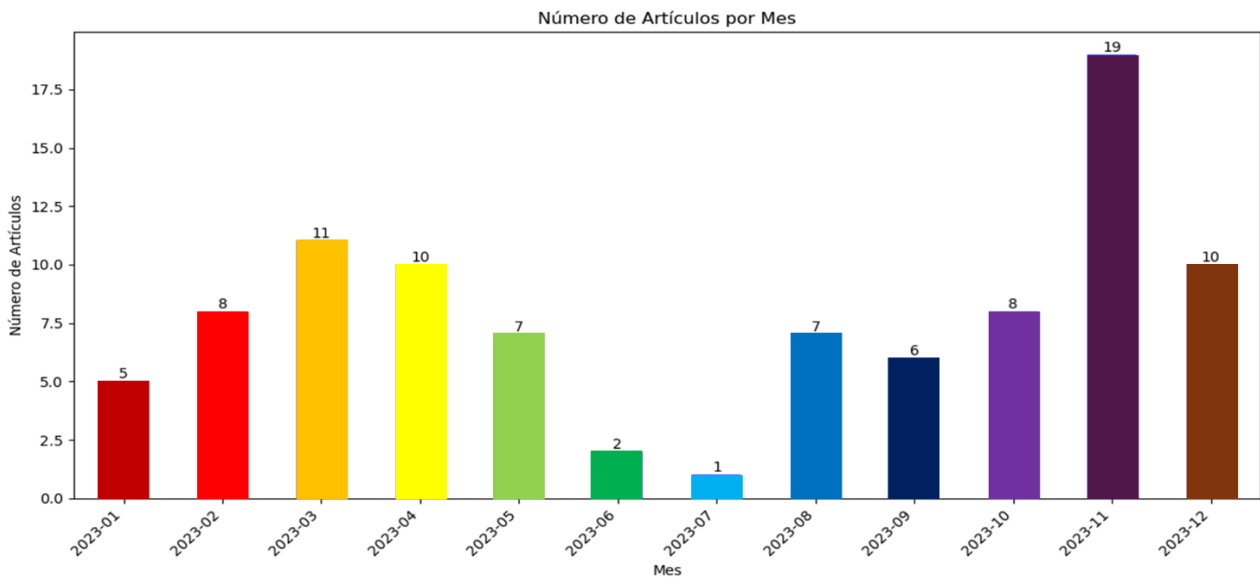
La ilustración 6 muestra la distribución del número de palabras en los artículos de opinión analizados sobre el tema de la paz en los principales periódicos de Bogotá. Se observa que los artículos varían significativamente en su longitud, desde un mínimo de 90 hasta un máximo de 1850 palabras. Esto refleja la diversidad en la profundidad y detalle con los que se aborda el tema de la paz en diferentes artículos.

El análisis del histograma destaca un pico significativo en la frecuencia de artículos que contienen entre 450 y 550 palabras. Esta concentración sugiere que una parte considerable de los artículos publicados en estos medios tiene una extensión moderada, lo cual puede estar relacionado con estilos editoriales que favorecen artículos concisos y directos. Después de este pico, se observa una disminución gradual en la frecuencia de artículos a medida que la longitud de los textos aumenta. No obstante, hay una ligera elevación en la frecuencia de artículos que cuentan con más de 1500 palabras. Esta tendencia puede indicar la presencia de artículos más detallados o analíticos que requieren una extensión adicional para desarrollar temas complejos o proporcionar análisis más profundos.



*Ilustración 8 Número de autores por cantidad de artículos de paz
Elaboración por autores*

La ilustración 8 muestra el análisis de la base de datos revela una notable participación de los ciudadanos en la expresión de sus opiniones en los periódicos de Bogotá, con un total de 75 autores identificados. De estos, 62 contribuyeron con un único artículo, 10 autores publicaron 2 artículos y 3 autores con más de 2 artículos. Esta diversidad de contribuciones no solo refleja la pluralidad de perspectivas sobre la paz, sino que también enriquece el análisis al incorporar múltiples puntos de vista que pueden ofrecer una comprensión más profunda de la perspectiva de paz en la sociedad bogotana en el año 2023.



*Ilustración 9 Número de artículos de paz por mes
Elaboración por autores*

En la ilustración 9 se observa que durante todo el año hay fluctuación en la cantidad de artículos publicados sobre paz; sin embargo, el tema está presente en los periódicos de Bogotá, lo que indica su importancia para la sociedad. El mes de noviembre de 2023 muestra el número más alto de artículos (19), lo que podría coincidir con eventos significativos relacionados con la paz, como negociaciones, conmemoraciones o declaraciones políticas. Esto sugiere un interés particular en el tema durante este mes. Por otro lado, los meses de junio (2) y julio (1) presentan la menor cantidad de artículos, lo que podría indicar un período de menor actividad en la discusión sobre la paz. Esto puede relacionarse con factores estacionales, como las vacaciones, que suelen disminuir la producción de contenido.

4.1.3. Preprocesamiento de datos

Para cada paso del preprocesamiento de los datos se requirió mantener la consistencia de los datos. Primero, se extrajo el archivo CSV generado en el paso anterior para obtener los datos procesados hasta ese punto. Luego, se definió una función específica para realizar la tarea deseada en ese paso, y que se explicarán a continuación. Una vez que se aplicó la función a los datos extraídos, se guardaron en un nuevo archivo CSV. Este ciclo se repitió en cada etapa del preprocesamiento, garantizando que los datos se procesaran de manera gradual y ordenada, y que cada transformación se aplicara de manera consistente en todo el conjunto de datos.

1. Reemplazo de letras con tildes por letras sin tildes

Durante este paso, se llevó a cabo el proceso de reemplazar todas las letras que contienen tildes por sus equivalentes sin tildes. Por ejemplo, la letra "á" se reemplazó con "a", "é" con "e", y así sucesivamente. Este reemplazo se realiza mediante un diccionario de mapeo de caracteres, este diccionario actúa como un conjunto de reglas predefinidas que especifican cómo transformar cada carácter especial con tilde en su forma estándar sin tilde. En el código, se emplea una expresión regular para encontrar las letras con tildes y luego se aplica la función de reemplazo utilizando el diccionario. Esto garantiza una estandarización coherente del texto, lo que es crucial para evitar posibles problemas de codificación o comparación entre palabras con y sin tildes. Esto es importante para estandarizar el texto y evitar posibles problemas de codificación o comparación entre palabras con y sin tildes. Sin embargo, es importante tener en cuenta que este proceso puede implicar la pérdida de información en algunos casos, especialmente cuando la tilde cambia el significado

2. Eliminación de caracteres no deseados

En esta etapa, se eliminaron todos los caracteres que no son letras del alfabeto y espacios en blanco del texto. Esto incluye signos de puntuación, números, símbolos especiales, etc. El objetivo es limpiar el texto y dejar solo los caracteres relevantes para el análisis. Para lograr esto, se definió un patrón de expresión regular `r'^[a-zA-ZáéíóúüñÁÉÍÓÚÜÑ\s]'`, el cual busca cualquier carácter que no sea una letra del alfabeto español (con o sin tilde) ni un espacio en blanco. Posteriormente, estos caracteres no deseados son reemplazados por espacios en blanco, dejando así un texto limpio y uniforme que facilita el procesamiento y análisis subsiguiente de los datos.

3. Convertir todo el texto a minúscula (normalización)

Aquí se convirtió todo el texto a minúsculas mediante el método `lower ()` de Python. Este paso es importante para la normalización del texto, ya que ayuda a garantizar que las palabras escritas de diferentes maneras (por ejemplo, "Hola" vs "hola") sean tratadas como iguales durante el análisis posterior.

4. Eliminación de *stopwords*

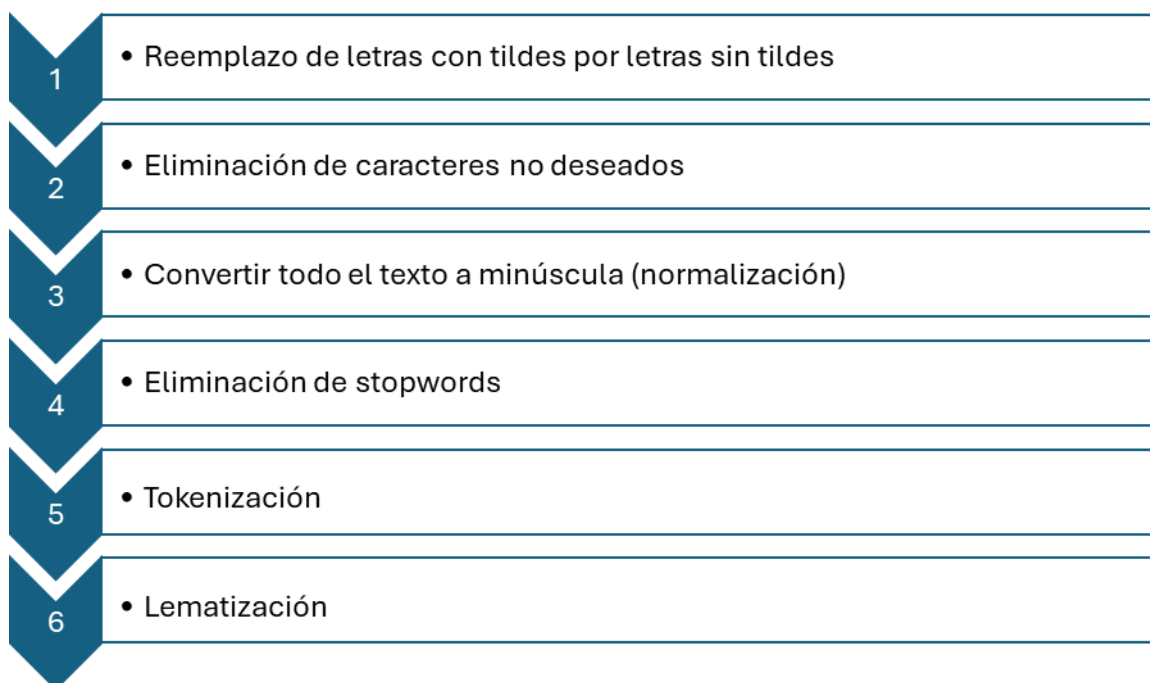
Durante este paso, se eliminaron las palabras comunes que no aportan un significado importante al texto, como artículos, preposiciones y conjunciones. Estas palabras, conocidas como *stopwords*, se eliminan para centrarse en las palabras clave que son más relevantes para el análisis. En esta implementación, se descargó la lista de *stopwords* en español de la librería Spacy para luego definir una función específica. Esta función recibe como entrada un texto y procede a dividirlo en palabras individuales. Luego, filtra aquellas palabras que no están presentes en la lista de *stopwords* descargada anteriormente. El resultado es un texto libre de *stopwords*, donde se han eliminado las palabras comunes que no aportan significado contextual al análisis.

5. Tokenización

La tokenización implica dividir el texto en unidades más pequeñas, llamadas *tokens*. En este caso, los tokens suelen ser palabras individuales, pero también podrían ser términos más grandes, dependiendo del contexto. La tokenización es esencial para el análisis del texto, ya que permite trabajar con unidades de texto más manejables y significativas. Para llevar a cabo esta implementación, fue esencial acceder al tokenizador de palabras provisto por spaCy. Se definió una función que, al recibir un texto como entrada, utiliza la función `word_tokenize` de spaCy para segmentar el texto en *tokens* individuales, los cuales representan palabras discretas. Como resultado, se genera una lista de palabras para cada texto ingresado, lo que facilita su posterior procesamiento y análisis.

6. Lematización

Finalmente, se realiza la lematización del texto, que consiste en reducir las palabras a su forma base o lema. Por ejemplo, las palabras "corriendo", "corrió" y "correrá" se reducirían todas a su forma base "correr". Esto ayuda a simplificar el análisis del texto al reducir la diversidad léxica y agrupar palabras relacionadas bajo una única forma. Para llevar a cabo esta implementación, se instaló el modelo de lenguaje `es_core_news_sm` de spaCy y se cargó para su uso. Luego, se leyeron los datos tokenizados de un archivo CSV, donde cada fila representa una lista de tokens correspondientes a palabras individuales en un texto. Utilizando el modelo cargado se tomó una lista de tokens, los unió en un texto y luego extrajo los lemas de cada palabra mediante el procesamiento con spaCy y estos lemas se almacenaron en una lista.



*Ilustración 10: Preprocesamiento de la base de datos
Elaboración por autores*

4.1.4. Identificación de palabras similares

Las palabras clave en este trabajo se consideran aquellas que presentan un alto grado de similitud con el término "paz". Este enfoque tiene como objetivo identificar los conceptos relacionados con la paz en el contexto de los bogotanos durante el año 2023. Al analizar estas palabras clave, se busca comprender mejor las percepciones, preocupaciones y temáticas que la población asocia con la paz, lo que permite una exploración más profunda de las dinámicas sociales en torno a este tema fundamental.

Después de lematizar el texto, se realizó un análisis de la frecuencia de las palabras mediante la creación de una tabla que muestra cuántas veces aparece cada palabra en el corpus de texto. Para lograr esto, se utilizó un contador que almacenó la frecuencia de las palabras. Este contador recorrió el archivo CSV que contenía el texto lematizado, actualizando las frecuencias de las palabras a medida que encontraba nuevas apariciones. Luego, las palabras se ordenaron de mayor a menor frecuencia.

Posteriormente, se cargó el modelo de lenguaje grande de spaCy en español y se definió la palabra "paz" como la palabra objetivo, se utilizó la parte del corpus correspondiente al año de interés, es decir del 2023, el cual se encontraba previamente procesado y tokenizado en lemas para calcular la similitud entre la palabra objetivo y cada palabra en el corpus. Esto se realizó mediante el método `similarity()` proporcionado por el objeto de documento de spaCy (`doc_objetivo`). Este método calcula la similitud del coseno entre dos vectores de palabras en el espacio vectorial. Posteriormente, se ordenaron las palabras relacionadas por su grado de similitud en orden descendente. Además, se generó la ilustración 11 para visualizar de manera más intuitiva estas

relaciones. Para esto, se utilizó la biblioteca WordCloud, que crea una nube de palabras donde el tamaño de cada palabra es proporcional a su aproximación con la palabra objetivo, “Paz”. La nube de palabras se construyó a partir de la tabla de palabras ordenadas previamente, y se mostró en una figura utilizando la biblioteca Matplotlib.



Ilustración 11: Palabras relacionadas con el término paz
Elaborada por autores

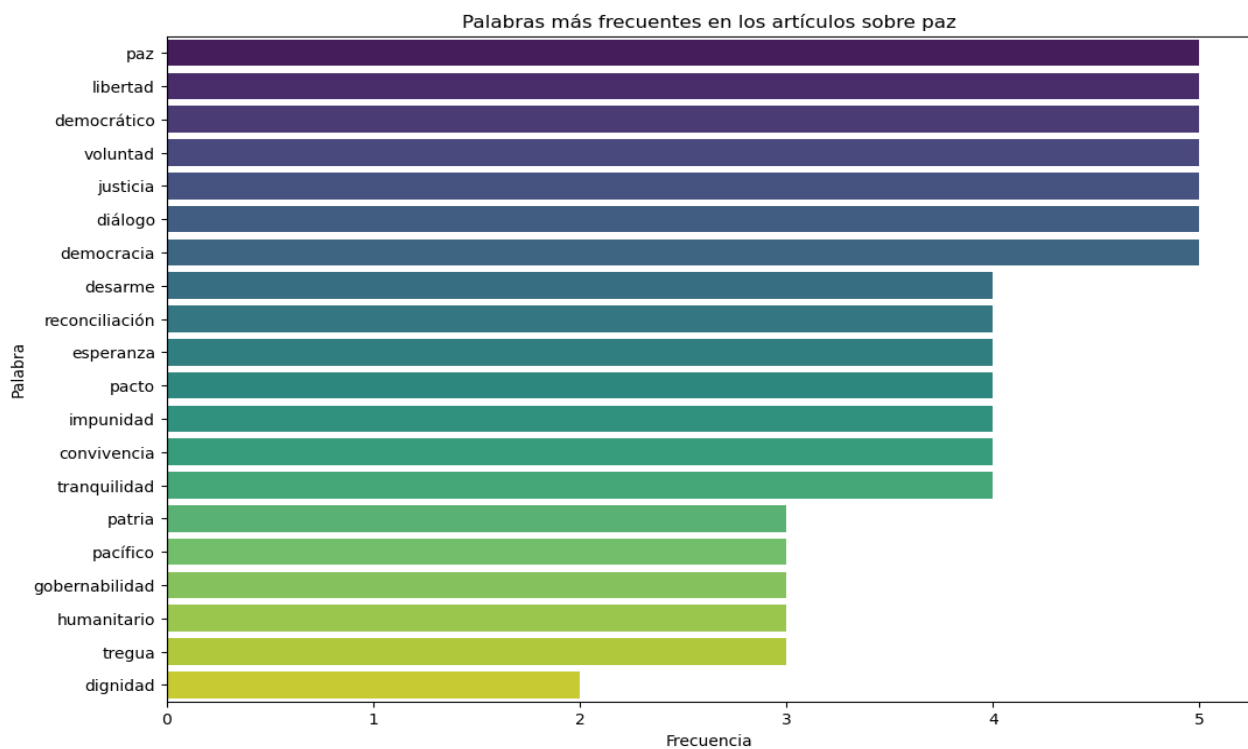
Lo descrito anteriormente se realizó para cada medio digital. Para ver las palabras asociadas con el concepto de 'paz' en cada medio digital, se creó la tabla 2 comparativa con el top 20 de palabras similares a 'paz' para cada uno. Esta tabla nos brinda un resumen conciso de las palabras clave más relevantes relacionadas con el tema de 'paz' en cada contexto periodístico:

TOP 20 SIMILITUD CON EL TERMINO PAZ					
	ESPECTADOR	LA REPUBLICA	LAS 2 ORILLAS	PORTAFOLIO	SEMANA
0	Paz	Paz	Paz	Paz	Paz
1	Reconciliación	Reconciliación	Reconciliación	Esperanza	Reconciliación
2	Esperanza	Fraternidad	Esperanza	Democracia	Esperanza
3	Pacífico	Democracia	Pacífico	Libertad	Pacífico
4	Serenidad	Libertad	Democracia	Justicia	Democracia
5	Democracia	Justicia	Libertad	Tranquilidad	Libertad
6	Libertad	Tranquilidad	Tregua	Calma	Tregua
7	Felicidad	Patria	Justicia	Patria	Justicia
8	Tregua	Comunicación	Solidaridad	Desarme	Tranquilidad
9	Justicia	Impunidad	Dignidad	Voluntad	Dignidad
10	Tranquilidad	Voluntad	Armisticio	Democrático	Patria
11	Solidaridad	Pacto	Desarme	Gobernabilidad	Desarme

12	Desarme	Democrático	Impunidad	Dialogo	Impunidad
13	Impunidad	Convivencia	Voluntad	Humanitario	Voluntad
14	Voluntad	Independencia	Pacto	Alegría	Pacto
15	Pacto	Pacificar	Democrático	Fortaleza	Democrático
16	Democrático	Igualdad	Convivencia	Confianza	Gobernabilidad
17	Gobernabilidad	Dialogo	Soberano	Guerra	Convivencia
18	Convivencia	Fortaleza	Dialogo	Humanidad	Dialogo
19	Independencia	Confianza	Humanitario	Liberación	Humanitario
20	Dialogo	Eternamente	Bienestar	Democráticamente	Pacificado

*Tabla 2 Cuadro comparativo top 20 palabras similares
Elaborada por autores*

Finalmente, se evaluaron las palabras consideradas en los periódicos mediante un gráfico de barras que muestra la frecuencia de cada término en el top 20 de palabras de los 5 periódicos seleccionados. Esta representación gráfica nos ayuda a identificar y comprender mejor los términos que la población asocia más estrechamente con el concepto de paz, proporcionando valiosa información sobre las percepciones de la sociedad en torno a la paz.



*Ilustración 12 Selección de términos relacionados con la Paz
Elaborada por autores*

En la ilustración 12 se observa que, los conceptos de libertad, democracia, voluntad, justicia y diálogo se encuentran consistentemente dentro del top 20 de los 5 periódicos seleccionados. Estos términos son indicativos de temas que están estrechamente relacionados con la percepción de paz por parte de la población y, por lo tanto, serían los indicados para ampliar la base de datos actual. Su presencia recurrente en los principales términos de cada periódico sugiere que estos

conceptos son de alta relevancia y preocupación para la sociedad en relación con la paz.

Además, se realizó un ejercicio similar con el concepto opuesto, 'guerra'. Los resultados de este análisis proporcionaron una comparación interesante para comprender mejor cómo se abordan y percibieron los temas de paz y guerra en la cobertura periodística.

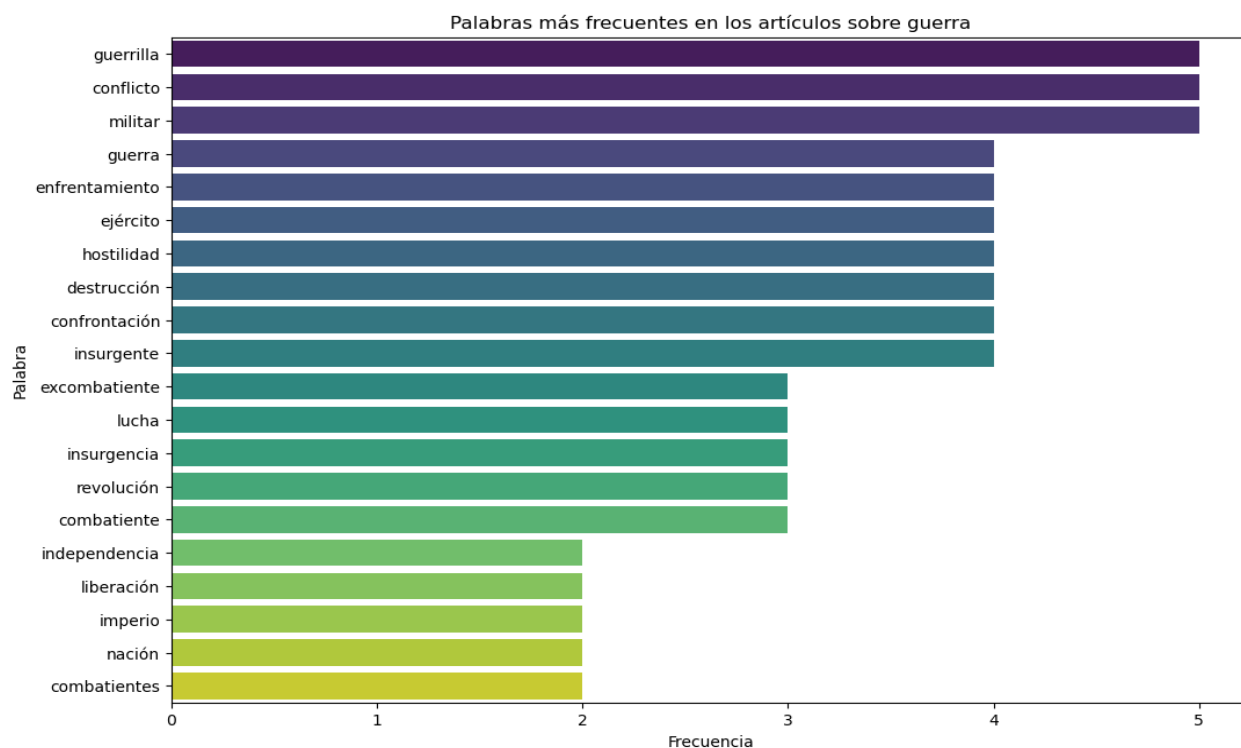


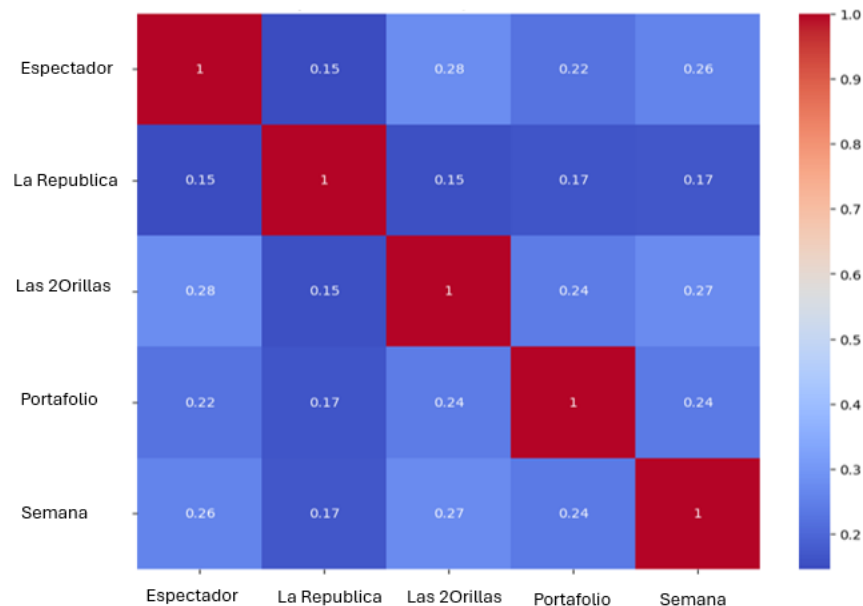
Ilustración 13 Selección de términos relacionados con la guerra
Elaborada por autores

Tras realizar el análisis de las palabras relacionadas con guerra, se determinó que las palabras identificadas no aportan información relevante al proyecto. En consecuencia, se ha decidido no implementarlas en el corpus ni en el análisis posterior. Esta decisión se basa en la evaluación de la pertinencia y utilidad de las palabras en relación con los objetivos del proyecto. En lugar de incluir términos que no contribuyen significativamente a nuestro análisis, nos centraremos en aquellos que son más relevantes y significativos para abordar adecuadamente el tema de estudio. De esta manera, garantizamos que nuestro análisis se centre en los aspectos más importantes y pertinentes relacionados con la percepción de paz en la población bogotana.

4.1.5. Similitud de Jaccard en el discurso de los periódicos

La similitud de Jaccard permite comparar la relación entre conjuntos de palabras asociadas con el concepto de "paz" en diferentes periódicos digitales, proporcionando una forma accesible de cuantificar hasta qué punto los distintos medios comparten una narrativa común sobre la paz o presentan enfoques divergentes. Esto facilita la identificación de patrones de similitud y diferencia en la cobertura mediática, ayudando a comunicar estos hallazgos de manera clara y comprensible.

A continuación, se muestra la ilustración 14 que muestra las similitudes de Jaccard entre las percepciones de paz de los diferentes periódicos digitales.



*Ilustración 14: Similitud de Jaccard entre periódicos
Elaborada por autores*

Los resultados de la similitud de Jaccard entre las palabras asociadas con el concepto de 'paz' en diferentes periódicos colombianos oscilan entre 0.15 y 0.28 (ver ilustración 14). Estos valores indican un nivel relativamente bajo de similitud entre los conjuntos de palabras utilizados por cada medio para describir la paz. Esto sugiere que los periódicos tienen enfoques y perspectivas diferentes sobre el concepto de paz. Aunque puede haber algunas palabras comunes, la mayoría de los términos utilizados por cada medio son distintos, reflejando sus prioridades editoriales y audiencias específicas.

Las2Orillas y El Espectador muestran una mayor similitud en sus narrativas de paz, posiblemente debido a enfoques editoriales similares en temas sociales y de derechos humanos. Mientras que entre La República y los otros medios, presentan una menor similitud en sus narrativas de paz, lo cual puede deberse a su enfoque más económico y empresarial, diferenciándose de los temas sociales y políticos que predominan en los otros periódicos.

Esta diversidad puede ser vista como una fortaleza, ya que proporciona múltiples perspectivas y enriquece el discurso público sobre un tema tan complejo y crucial. La variabilidad en las narrativas mediáticas subraya la importancia de considerar estas diferentes perspectivas para obtener una comprensión más completa y matizada de lo que significa la paz en el contexto colombiano.

4.1.6. Análisis de Palabras Clave con YAKE (Yet Another Keyword Extractor)

Para confrontar los resultados obtenidos con el análisis anterior, se utilizó la herramienta YAKE para extraer las palabras y frases más relevantes de los artículos relacionados con la paz del 2023 (El mismo corpus del análisis anterior). Primero, se cargaron los artículos en formato CSV, y el texto fue preprocesado para convertirlo a mayúsculas, asegurando la uniformidad en el análisis y evitando sesgos provocados por las diferencias en la capitalización de las palabras. A continuación, se configuró el extractor YAKE para trabajar en idioma español, con un tamaño de n-gramas de hasta tres palabras y un umbral de deduplicación de 0.9, lo que permitió eliminar las palabras repetidas y mantener aquellas que realmente reflejaban los temas más relevantes. Además, se definió una lista de términos a excluir, tales como "GUSTAVO", "PETRO", "PRESIDENTE", Y "COLOMBIA", con el fin de evitar que términos demasiado específicos o asociados a figuras políticas distorsionaran el análisis. De esta manera, se pudo enfocar la extracción de palabras clave en conceptos más relevantes y representativos de los temas con los que los bogotanos relacionaron la paz en 2023.

Una vez preprocesado el texto y configurado el extractor YAKE, se procedió a aplicar la herramienta a cada artículo del corpus. Para cada artículo, se extrajeron las 10 palabras clave más relevantes, garantizando que las palabras seleccionadas fueran las que mejor representaran los temas tratados en el texto. Este paso permitió obtener un conjunto de palabras clave específicas para cada artículo, lo que proporcionó una visión detallada de los temas más destacados en los distintos textos. Posteriormente, con el objetivo de identificar las palabras clave más recurrentes en el conjunto total de artículos, se consolidaron todas las palabras clave extraídas en una lista general. Para ello, se realizó un proceso de agregación en el que se contaron las frecuencias de aparición de cada palabra clave. Este paso permitió eliminar cualquier variabilidad individual de los artículos y destacar las palabras que se presentaron con mayor frecuencia a través del conjunto completo de documentos.

Este conjunto de palabras clave ofreció una representación más clara de los temas predominantes en los artículos, destacando conceptos clave como "PAZ", "GOBIERNO", "DEMOCRACIA", "JUSTICIA", y "LIBERTAD". Sin embargo, también se obtuvo una cantidad significativa de términos que representaron ruido para los objetivos del proyecto. Algunos de estos términos, como "CONMEBOL TOMA RADICAL", "EMAIL PROTECTED" y "CASTELLANA BOGOTÁ", no estuvieron directamente relacionados con el tema de la paz en 2023. Estos resultados fueron considerados irrelevantes y, por lo tanto, se eliminaron mediante un proceso de filtrado adicional. Esto permitió asegurar que las palabras clave extraídas se centraran únicamente en aquellos conceptos más representativos y relevantes para la discusión sobre la percepción de paz en Colombia.

En comparación con el análisis previo los resultados obtenidos mediante el extractor YAKE mostraron una concordancia notable en los temas más recurrentes, tanto en el análisis anterior como en este, los términos "PAZ", "DEMOCRACIA", "JUSTICIA" y "LIBERTAD" aparecieron entre los más relevantes. Esta coincidencia reforzó la idea de que estos términos son fundamentales para entender como es la percepción de paz de los bogotanos en el 2023.

4.1.7. Creación del corpus de trabajo

Durante toda la investigación, las palabras **Libertad**, **Democracia** y **Justicia** se destacan entre los resultados obtenidos, reflejando su relevancia para los habitantes de Bogotá. Este hallazgo sugiere que la paz, para ellos, trasciende la ausencia de conflicto armado y se entiende como un estado integral vinculado a la equidad, la reparación a las víctimas y la reconstrucción del tejido social. La libertad se asocia con la capacidad de los ciudadanos para expresar sus opiniones y participar activamente en la sociedad, mientras que la democracia implica un sistema donde se respetan los derechos de todos y se fomenta la participación ciudadana en la toma de decisiones. Por otro lado, la justicia se vincula con la equidad y la resolución de conflictos de manera pacífica, elementos esenciales para lograr y mantener un entorno de paz duradera. Juntas, estas palabras clave ayudan a construir un marco de referencia para entender cómo los bogotanos conceptualizan la paz en el contexto de su realidad cotidiana y sus aspiraciones colectivas. Este conjunto de términos sugiere una visión compartida donde la paz se asocia estrechamente con valores fundamentales y procesos sociales inclusivos y equitativos. Podemos ver la visión de cada periódico de la siguiente manera:

- El Espectador: Destaca términos como 'serenidad', 'felicidad' y 'solidaridad', sugiriendo una visión de paz que incluye bienestar emocional y cohesión social.
- La República: Este medio incluye palabras como 'fraternidad', 'patria', e 'igualdad', indicando un enfoque hacia la paz que enfatiza la unidad nacional y la equidad.
- Las2Orillas: Este periódico resalta palabras como 'tregua', 'dignidad', y 'convivencia', reflejando un enfoque en la reducción del conflicto, la promoción de la dignidad humana y la coexistencia pacífica.
- Portafolio: Con términos como 'esperanza', 'democracia', 'libertad', y 'justicia', Portafolio mantiene un enfoque similar a otros medios, pero también incluye palabras como 'desarme' y 'gobernabilidad', sugiriendo una preocupación por el desarme y la estabilidad política.
- Semana: Además de compartir muchos términos comunes, Semana incorpora palabras como 'pacificado', 'gobernabilidad' y 'humanitario', lo cual indica un enfoque en la implementación práctica de la paz y las dimensiones humanitarias.

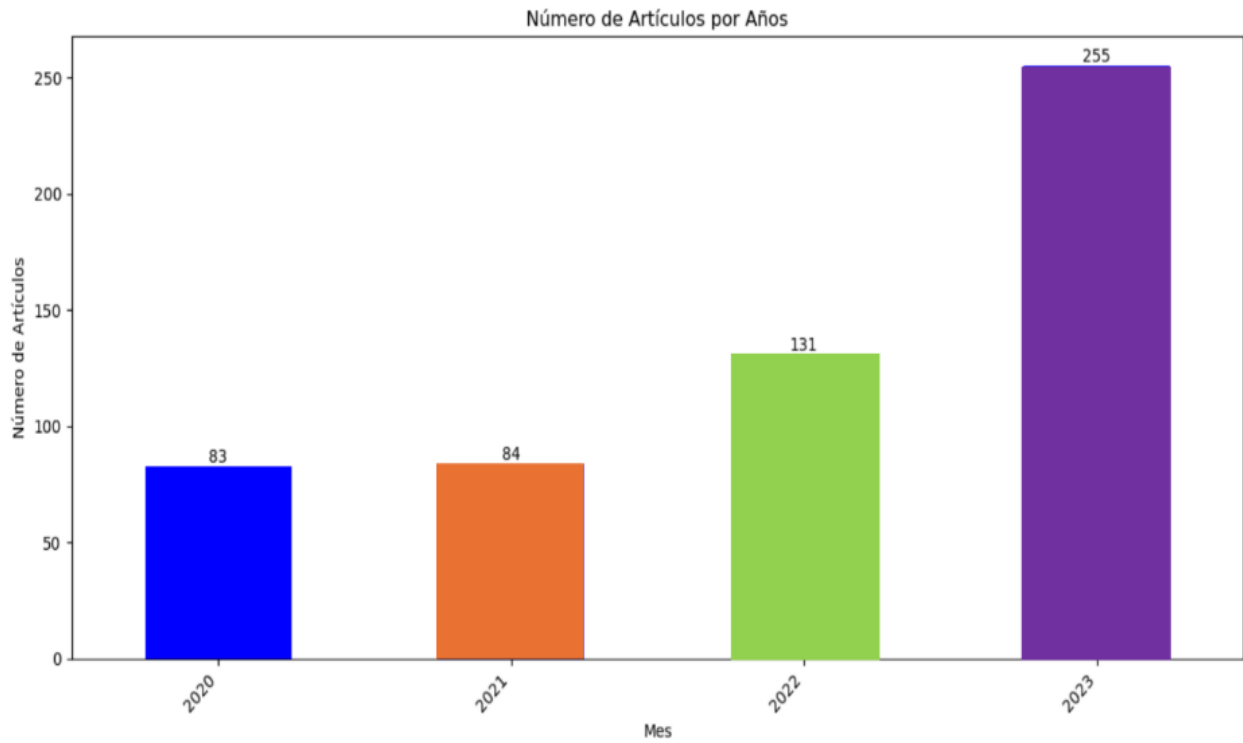
Partiendo de las nuevas palabras identificadas, se realizaron nuevamente los procesos de búsqueda, web scrapping y preprocesamiento de la base de datos con el fin de completar el corpus del trabajo. Este proceso garantiza que nuestro corpus refleje con precisión los temas relacionados con la percepción de paz. Este corpus está compuesto por artículos de opinión que contienen las palabras "Paz", "Libertad", "Democracia" y "Justicia", y proviene de los periódicos seleccionados: El Espectador, Portafolio, La República, Semana y Las 2 Orillas, en el periodo desde el 1 de enero de 2020 hasta el 31 de diciembre de 2023. La información extraída incluyó el nombre del periódico, el título del artículo, el autor, la fecha de publicación, el número de palabras y el texto completo del artículo. A continuación, se presenta la conformación de este corpus.

	NOMBRE DEL PERIÓDICO	NÚMEROS DE ARTÍCULOS
1	EL ESPECTADOR	173
2	PORTAFOLIO	93
3	LA REPÚBLICA	77
4	SEMANA	95
5	LAS 2 ORILLAS	115
	TOTAL	553

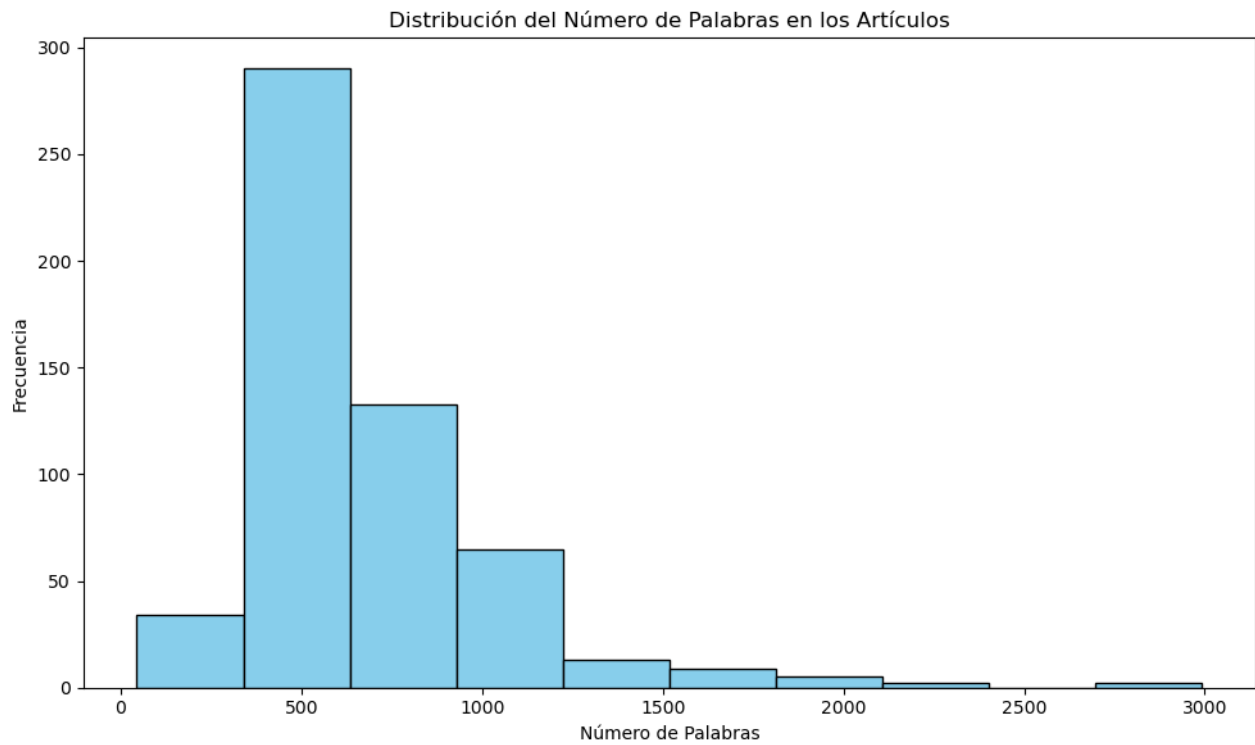
*Tabla 3 Composición de la base de datos final
Elaborada por autores*

La tabla 3 resume la cantidad de artículos recopilados de cada periódico durante el periodo de estudio, totalizando 553 artículos. Este corpus amplio y diverso nos permitirá contar con la mayor cantidad de matices de la percepción de paz para los bogotanos. A continuación, se presenta un análisis de la distribución de estos artículos por año, lo cual nos permitirá identificar tendencias y patrones en la cobertura de estos temas a lo largo del tiempo.

Al analizar la distribución de artículos por año, observamos un aumento en el número de artículos desde 2020 hasta 2023. Este patrón sugiere una creciente atención y relevancia de los temas relacionados con la "Paz" en los medios de comunicación bogotanos, lo cual podría indicar una mayor preocupación por expresar las opiniones de la sociedad ante los eventos que motivan la percepción de paz. La base de datos creada es un recurso adecuado para llevar a cabo la investigación propuesta, tanto en cantidad como en representatividad, ya que refleja con precisión la diversidad y riqueza de las discusiones sobre la paz en Bogotá durante este periodo.



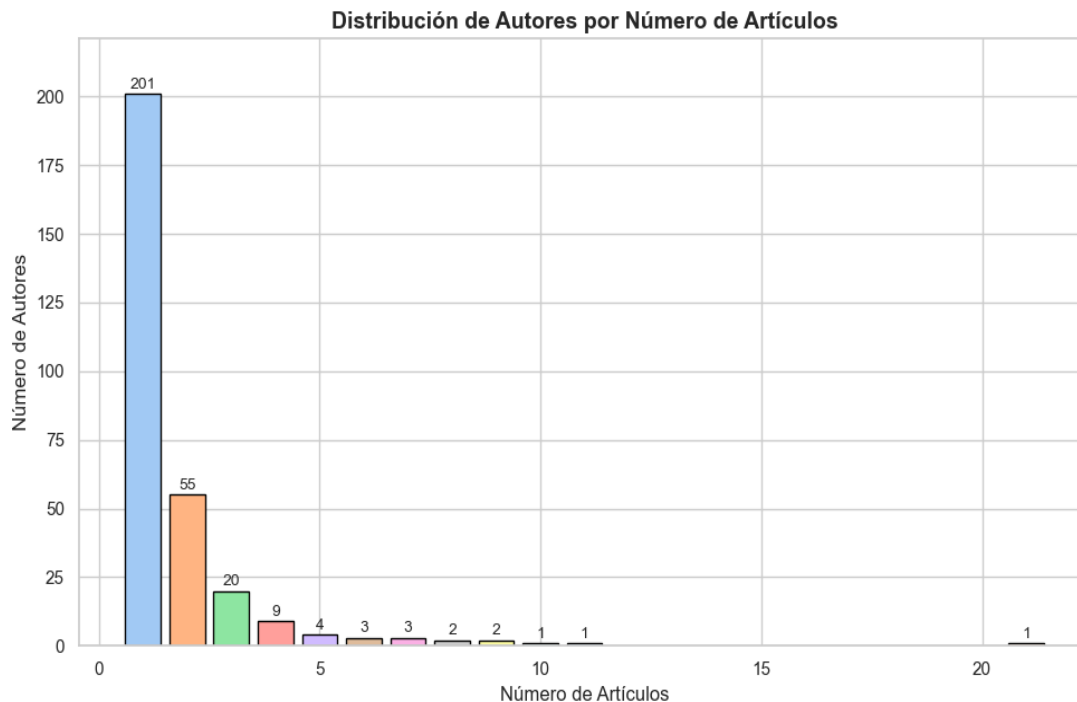
*Ilustración 15 Número de artículos por año de la base de datos
Elaboración por autores*



*Ilustración 16 Distribución del número de palabras en los artículos
Elaboración por autores*

Siguiendo con la tendencia, antes descrita cuando solo se contaban con los artículos con el término paz del 2023, observamos que la longitud de los artículos se mantiene con una forma asimétrica hacia la izquierda, es decir, el pico de la ilustración 16 se mantiene en la frecuencia de que contienen entre 450 y 550 palabras. Esta concentración sugiere que una parte considerable de los artículos publicados en estos medios tiene una extensión moderada, lo cual puede estar relacionado con estilos editoriales que favorecen artículos concisos y directos. Después de este pico, se observa una disminución gradual en la frecuencia de artículos a medida que la longitud de los textos aumenta.

Esta variabilidad en la longitud de los artículos no solo enriquece el corpus, sino que también permite un análisis más matizado y profundo de la percepción de paz en Bogotá. Al incluir tanto artículos breves, con 44 palabras, como extensos, que alcanzan hasta 2993 palabras, se pueden captar diferentes enfoques y niveles de profundidad en el tratamiento de los temas. Los artículos más cortos pueden ofrecer una visión rápida y directa de opiniones y eventos, mientras que los más largos tienen la capacidad de explorar contextos históricos, sociales y políticos, proporcionando un análisis más completo y detallado. Esta diversidad en la extensión de los artículos también refleja la pluralidad de voces y perspectivas presentes en el debate sobre la paz, lo que resulta fundamental para entender cómo diferentes segmentos de la sociedad bogotana perciben y discuten este tema crucial.



*Ilustración 17 Número de autores por cantidad de artículos
Elaboración por autores*

La ilustración 17 permite analizar el número de artículos por autor revela que la distribución de contribuciones en el corpus estudiado se mantiene con la tendencia antes descrita, cuando solo se contaban con los artículos con el término paz del 2023. Observamos que se identificaron 201

autores que publicaron un único artículo, lo que sugiere una amplia variedad de voces y perspectivas en la discusión sobre los temas relacionados con la "Paz". Además, 55 autores contribuyeron con dos artículos, 20 autores escribieron tres, y otros 9 autores publicaron cuatro. Este patrón continúa disminuyendo hasta alcanzar contribuciones más bajas. La recopilación de los artículos se llevó a cabo desde el año 2020 hasta 2023, un periodo que ha permitido capturar una variedad significativa de contribuciones, sin embargo, debido a la duración de este periodo y la naturaleza del trabajo periodístico, muchos autores se repiten, lo que indica que ciertos escritores han estado activos en la expresión de estos temas.

Es relevante señalar que algunos artículos se presentan bajo el nombre del periódico debido a la falta de información sobre el autor. Esta circunstancia limita la representación de voces individuales en el corpus, aunque la cantidad de artículos contribuye a la riqueza del análisis. La diversidad en la cantidad de artículos por autor resalta la pluralidad de opiniones y enfoques, lo que enriquece la expresión de la percepción de paz en Bogotá en los periódicos y proporciona una base sólida para la investigación.

4.2 OBJETIVO 2

Desarrollar un modelo basado en procesamiento de lenguaje natural y técnicas de análisis de sentimiento que permita cuantificar la percepción de paz a partir de la categorización de expresiones relacionadas al tema en las plataformas digitales.

De la base de datos recopilada, se decidió utilizar los artículos del año 2023, que suman un total de 255, para la evaluación de la percepción de paz en Bogotá en ese año, mientras que los datos correspondientes a los años 2020 al 2022, que consisten en 292 artículos, se emplearon para el entrenamiento del modelo. Los artículos del conjunto de entrenamiento se etiquetaron manualmente de acuerdo con su sentimiento (negativo, positivo o neutral) con respecto a la percepción de paz de los bogotanos, partiendo de la definición de paz como libertad, democracia y justicia. Esta separación entre los conjuntos de datos permite que el modelo se entrene con información relevante, lo que garantiza una mejor comprensión de las dinámicas y tendencias en la discusión sobre la paz.

4.2.1 Análisis de sentimiento

4.2.1.1 Modelo no supervisado

Se intentó aplicar un modelo de análisis de sentimiento no supervisado utilizando la librería TextBlob, que permite realizar análisis de sentimientos a partir de las características léxicas del texto. En este caso, se implementó una función denominada `analizar_sentimiento` que toma una lista de lemas (palabras clave procesadas del texto) y realiza un análisis de polaridad para determinar el sentimiento. La polaridad es un valor numérico que varía entre -1 y 1, donde valores positivos indican sentimientos positivos, valores negativos indican sentimientos negativos y valores cercanos a cero corresponden a sentimientos neutros. El proceso de análisis consistió en lo siguiente: primero, se unieron los lemas en una cadena de texto, luego se creó un objeto

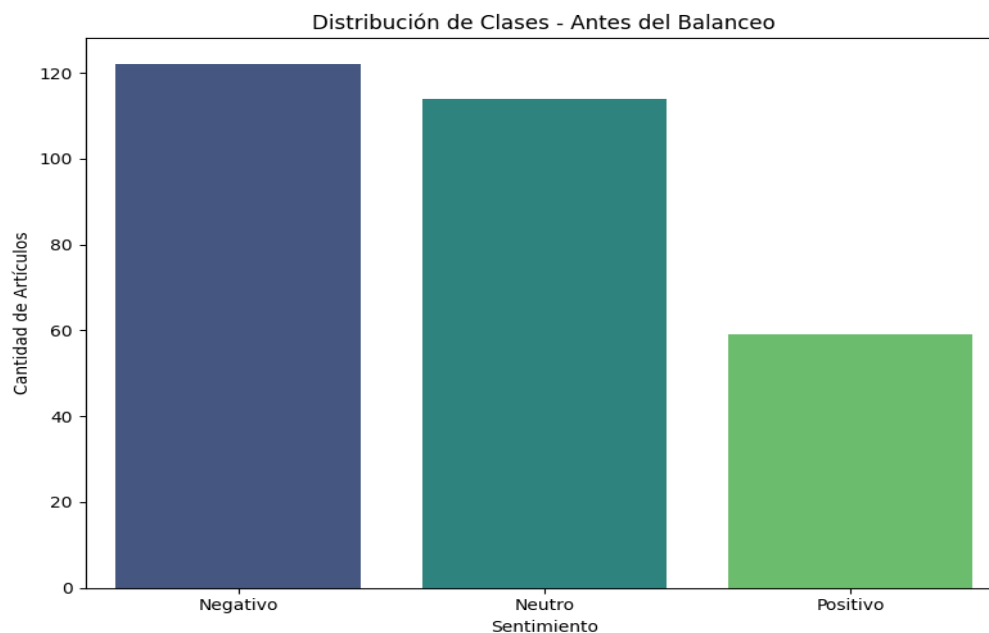
TextBlob a partir de esta cadena. A través de la propiedad `sentiment.polarity`, se obtuvo la polaridad del texto. Según el valor de esta polaridad, se asignó una etiqueta de sentimiento: 'Positivo' para valores superiores a 0, 'Negativo' para valores inferiores a -0.1 y 'Neutro' para valores intermedios. Finalmente, esta función se aplicó a cada texto del conjunto de datos de entrenamiento, almacenando los resultados en una nueva columna llamada `sentimiento_predicho`. A pesar de ser un enfoque no supervisado, esta técnica nos permitió comprender mejor la estructura y las características del conjunto de datos, brindando una visión más clara sobre la distribución de sentimientos en los textos. Esta exploración inicial, aunque sencilla, proporcionó información valiosa sobre los patrones subyacentes en la data y facilitó la identificación de posibles ajustes y mejoras para el análisis más profundo con modelos supervisados.

4.2.1.2 Modelo Supervisado

Para analizar el sentimiento presente en el corpus de texto, se empleó un enfoque basado en modelos de lenguaje preentrenados, específicamente utilizando BERT (Bidirectional Encoder Representations from Transformers) junto con la biblioteca Hugging Face's Transformers en Python. Inicialmente, se cargó el modelo preentrenado BERT, denominado 'dccuchile/bert-base-spanish-wwm-cased', el cual ha sido entrenado en una amplia variedad de textos en español. A su vez, se empleó un tokenizador específico para este modelo, encargado de dividir las frases en tokens compatibles con la arquitectura de BERT, ajustándolas a una longitud máxima de 512 tokens e incorporando los tokens especiales requeridos. La selección de esta longitud máxima responde a las limitaciones inherentes de la arquitectura BERT, ya que, admite secuencias de hasta 512 tokens. Este límite permite un balance entre la representación del contenido y la eficiencia computacional del modelo. Una vez que los textos fueron tokenizados, se procesaron en lotes que alimentaron al modelo BERT, obteniendo así una representación contextualizada de las palabras en cada frase. A continuación, el modelo aplicó una capa de clasificación para asignar una probabilidad a cada una de las tres categorías de sentimiento (negativo, neutral y positivo), utilizando una función de activación softmax para normalizar estas probabilidades. Las etiquetas de sentimiento finales se asignaron en función de la categoría con mayor probabilidad.

A pesar de este proceso, los resultados obtenidos no fueron consistentes. En varias ejecuciones, las predicciones variaban significativamente, lo que implicaba que no siempre se lograba identificar correctamente las tres categorías de sentimiento. Esta variabilidad en los resultados reveló una inestabilidad en el desempeño del modelo. Dado que se aplicó el modelo base sin ajustes específicos al conjunto de datos utilizado, esto sugiere que BERT no fue capaz de capturar adecuadamente las sutilezas del tema en análisis. Ante esta situación, se decidió llevar a cabo un ajuste fino del modelo. Como primer paso, se realizó un análisis de la conformación de los datos de entrenamiento. Este análisis reveló una distribución desbalanceada entre las clases: 122 artículos con sentimiento negativo, 114 con sentimiento neutral y solo 59 con sentimiento positivo, como se ilustra en la gráfica siguiente. Esta disparidad indicaba que el modelo podría estar sesgado hacia las clases mayoritarias, afectando tanto su precisión como su capacidad de generalización. Para mitigar este problema, se optó por balancear las clases, ajustando el conjunto

de entrenamiento de modo que el modelo pudiera aprender de manera equitativa las tres categorías de sentimiento, lo que mejoraría su desempeño en la evaluación posterior.

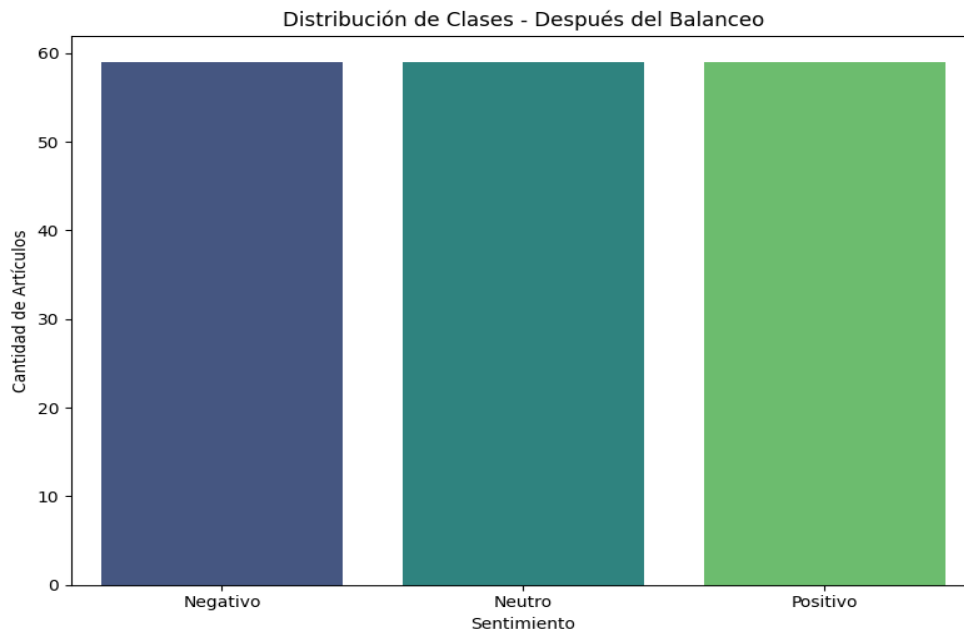


*Ilustración 18: Distribución de clases en base de entrenamiento antes del balanceo
Elaboración por autores*

Para asegurar que el modelo de entrenamiento sea robusto y no sesgado hacia ninguna clase en particular, se realizó un balance de clases en los datos utilizando la técnica de submuestreo. Esto implicó reducir la cantidad de artículos en las clases más representadas para que estuvieran en proporciones similares a las clases menos representadas en el conjunto de datos de entrenamiento. Al submuestrear las clases más prevalentes, logramos una distribución equilibrada que mejora la capacidad del modelo para identificar y clasificar correctamente las distintas percepciones de paz expresadas en los artículos. Este proceso es crucial para garantizar que el análisis y las predicciones reflejen de manera justa y precisa las diferentes perspectivas sobre la paz en Bogotá. La principal herramienta empleada fue pandas, que facilitó la manipulación y manejo de los datos en formato de DataFrame. A través de la función `groupby()`, se organizó el conjunto de datos según las categorías de sentimiento (negativo, neutral y positivo), se utilizó la biblioteca scikit-learn, específicamente la función `resample()`, que permitió reducir el número de ejemplos de las clases mayoritarias (negativo y neutral) de forma aleatoria. Esta función garantizó que las muestras seleccionadas fueran representativas de la clase original, eliminando la posibilidad de sesgos que podrían haberse introducido por una selección no aleatoria.

La lógica propuesta para balancear las clases consistió en igualar la cantidad de artículos en todas las categorías de sentimiento. Dado que la clase positiva tenía 59 artículos, se estableció este número como referencia, reduciendo las clases mayoritarias (negativo y neutral) a este valor. El submuestreo se realizó sin reemplazo, lo que significó que, de las clases mayoritarias, se extrajeron aleatoriamente solo 59 artículos sin duplicar ningún dato. Una vez ejecutado este

proceso, se verificó la nueva distribución utilizando la función `value_counts()` de pandas, lo que reveló que cada clase contenía ahora 59 artículos, sumando un total de 177 artículos en el conjunto de entrenamiento. Este balance de clases contribuyó significativamente a mejorar la capacidad del modelo para aprender y clasificar las tres categorías de manera equitativa, evitando sesgos hacia las clases mayoritarias y promoviendo una clasificación más justa y precisa de las diferentes percepciones sobre la paz expresadas en los artículos.



*Ilustración 19: Distribución de clases en base de entrenamiento después del balanceo
Elaboración por autor*

El modelo BERT para análisis de sentimiento clasifica el texto en función de las emociones y actitudes expresadas por las palabras. Sin embargo, para este estudio en particular, se requiere una evaluación más específica: determinar si el texto se alinea con los sentimientos y perspectivas a favor o en contra de la percepción de paz de los bogotanos, es decir, en relación con la democracia, la justicia y la libertad. En lugar de simplemente categorizar el texto como positivo, neutral o negativo, es crucial identificar si el contenido refleja una actitud favorable hacia estos aspectos. Para lograr esto, el análisis de sentimiento debe adaptarse a través de un proceso de refinamiento, conocido como "fine-tuning" o ajuste fino. Este proceso consiste en tomar un modelo preentrenado y adaptarlo a una tarea específica mediante el entrenamiento adicional con un conjunto de datos relevante. Esta técnica mejora el rendimiento del modelo en tareas particulares al ajustar sus parámetros de manera más precisa, logrando así una mayor precisión y eficacia en los resultados.

Para el refinamiento del modelo se prepararon y dividieron los datos en conjuntos de entrenamiento y validación, asegurando que el modelo pueda aprender de un 80% de los datos mientras se evalúa en un 20%. Esta proporción se eligió para maximizar la cantidad de información disponible durante el entrenamiento, lo que permite al modelo captar patrones más

representativos, mientras que el 20% restante se reservó para la evaluación, garantizando así una validación efectiva del rendimiento del modelo. Se utilizó la información correspondiente a los años 2020 al 2022, cuyos artículos fueron etiquetados manualmente mediante una revisión cuidadosa. Durante este proceso, se asignaron categorías específicas a los textos en función de su alineación con la percepción de paz de los bogotanos. A continuación, se creó un Dataset que organizó la información y se utilizó un DataLoader para alimentar los datos al modelo en lotes. Se cargó el modelo BERT preentrenado y se define un optimizador que ajustará sus pesos durante el entrenamiento, complementado por un scheduler que controla la tasa de aprendizaje. Durante el entrenamiento, se itera a través del DataLoader, pasando los datos al modelo, calculando la pérdida y ajustando parámetros, mientras se evaluó su desempeño en el conjunto de validación. Finalmente, se guarda el modelo refinado y el tokenizador, lo que permite su uso posterior para clasificar nuevos datos.

Antes de realizar el fine-tuning, se observó que el modelo BERT operaba de manera más efectiva cuando trabajaba a partir del contexto completo de las ideas expresadas en los textos, en lugar de depender únicamente de un texto preprocesado y lematizado. Esta observación surgió al notar que la eliminación de información contextual importante durante el preprocesamiento podría estar limitando la capacidad del modelo para captar las relaciones semánticas entre las palabras y las frases. Por lo tanto, se decidió abandonar el enfoque de trabajar con textos preprocesados y lematizados, optando por utilizar el texto completo en su forma original. Este cambio permitió que BERT pudiera aprovechar mejor el contexto global de las ideas expresadas.

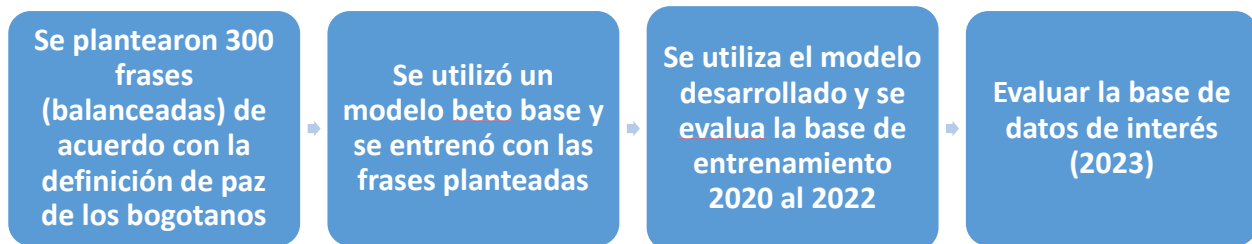
La etapa de fine-tuning se realizó mediante tres técnicas diferentes para mejorar la estabilidad y precisión del modelo. Primero, se balancearon los textos utilizando submuestreo, reduciendo la cantidad de ejemplos de las clases más representadas para igualar el número de instancias de cada clase. Segundo, se aplicó un balanceo con sobremuestreo, aumentando artificialmente el número de ejemplos de las clases menos representadas mediante la técnica de sinónimos, reemplazando ciertas palabras de los textos por sus equivalentes semánticos. Por último, se entrenó al modelo utilizando únicamente los textos con sentimientos extremos (positivo y negativo), instruyendo al modelo a reconocer tres clases, a pesar de que solo se le proporcionaron ejemplos de dos. Sin embargo, ninguna de estas estrategias resultó ser efectiva. A pesar de los esfuerzos por ajustar los hiperparámetros para lograr un mejor rendimiento, todas las técnicas de fine-tuning produjeron un accuracy inferior al 35%, lo que sugiere que el modelo no fue capaz de capturar adecuadamente las sutilezas del corpus ni reflejar con precisión las percepciones de sentimiento en los textos analizado.

Después de los resultados insatisfactorios en las pruebas previas, se decidió utilizar el modelo preentrenado "nlptown/bert-base-multilingual-uncased-sentiment", diseñado para predecir sentimientos, se cargó el modelo y se dividieron los artículos en fragmentos basados en los puntos seguidos, asegurando que cada uno no excediera los 512 tokens, límite que BERT puede procesar. Esta segmentación permitió conservar el contexto completo de los textos sin perder información relevante. Cada fragmento fue clasificado en sentimientos negativos, neutros o positivos, según los puntajes obtenidos del modelo. Posteriormente, se calculó un resultado de sentimiento para

cada artículo promediando los puntajes de todos sus fragmentos, lo que permitió asignar un único valor de sentimiento por texto completo. Luego, se compararon estas predicciones con las etiquetas esperadas previamente definidas de manera manual. Para medir el rendimiento del modelo, se calcularon métricas como precisión, recall y F1-score, lo que permitió evaluar su capacidad para identificar correctamente los distintos sentimientos en los textos. El modelo alcanzó una precisión global del 42%, mejorando los intentos anteriores, pero con un desempeño limitado, especialmente en la clasificación de sentimientos positivos. Aunque el modelo identificó de forma razonable los sentimientos negativos, los resultados para las clases neutra y positiva fueron insuficientes.

4.2.1.3 Modelo final

Para mejorar el desempeño de los modelos, se decidió cambiar la estrategia de entrenamiento en la fase de fine-tuning, siguiendo el esquema detallado en la imagen



*Ilustración 20 Proceso final modelo análisis de sentimiento
Elaboración por autores*

En este enfoque, se plantearon 336 frases diseñadas de manera estratégica para reflejar tres tipos de sentimientos: 112 positivas, 112 neutras y 112 negativas. Estas frases fueron cuidadosamente alineadas con la definición de paz de los bogotanos, basándose en los conceptos clave de paz, libertad, democracia y justicia.

Las frases positivas se enfocan en resaltar avances, mejoras o logros relacionados con la paz, la justicia, la democracia y la libertad. Se utilizan palabras que generan optimismo, confianza y orgullo. Por ejemplo: *“En un esfuerzo por fortalecer la democracia, la alcaldía de Bogotá ha implementado nuevos mecanismos para incentivar la participación ciudadana en decisiones clave. A través de plataformas digitales y cabildos abiertos, los ciudadanos ahora tienen un espacio para expresar sus preocupaciones y propuestas. Esta iniciativa busca fomentar un diálogo abierto y transparente que refuerce los valores democráticos en la capital.”*

Las frases neutras describen hechos o situaciones de manera objetiva, sin cargar emocionalmente el mensaje. Son frases que informan o presentan la situación sin emitir juicios valorativos o emociones, manteniendo un tono imparcial. Por ejemplo: *“A pesar de los esfuerzos por mejorar el acceso a la justicia en Bogotá, las comunidades más vulnerables aún enfrentan grandes barreras. El sistema judicial sigue siendo lento e ineficiente, con pocos avances significativos en la igualdad de acceso. La paz basada en la justicia solo será posible cuando estas barreras sean*

eliminadas.”

Las frases negativas ponen énfasis en los problemas, retos o retrocesos que afectan negativamente estos mismos conceptos, resaltando conflictos o injusticias que afectan la paz. Se recurre a un lenguaje que evoca frustración, crítica o preocupación. Por ejemplo: *“Recientes restricciones a la libertad de expresión en Bogotá han generado preocupación entre diversos sectores. Si bien algunos sostienen que estas medidas buscan mantener el orden público, otros consideran que podrían socavar la democracia y la justicia. Es un tema que requiere más análisis antes de emitir un juicio definitivo.”*

El proceso de entrenamiento del modelo se llevó a cabo utilizando el modelo preentrenado dccuchile/bert-base-spanish-wwm-cased. Como primer paso, las frases desarrolladas se prepararon asignando etiquetas numéricas a los sentimientos identificados en los textos (Negativo: 0, Neutro: 1, Positivo: 2). Para asegurar una representación balanceada de cada clase, se aplicaron técnicas de submuestreo, equilibrando así las instancias de cada etiqueta y evitando posibles sesgos de clase que pudieran afectar la capacidad de generalización del modelo. Con los datos balanceados, se procedió a la tokenización de las frases mediante el tokenizador correspondiente a BERT, adaptado para procesar texto en español. Posteriormente, los datos tokenizados se dividieron en dos conjuntos: el conjunto de entrenamiento (80%) y el conjunto de validación (20%). Esta división permitió evaluar el rendimiento del modelo en datos no vistos y fue crucial para detectar y evitar problemas de sobreajuste.

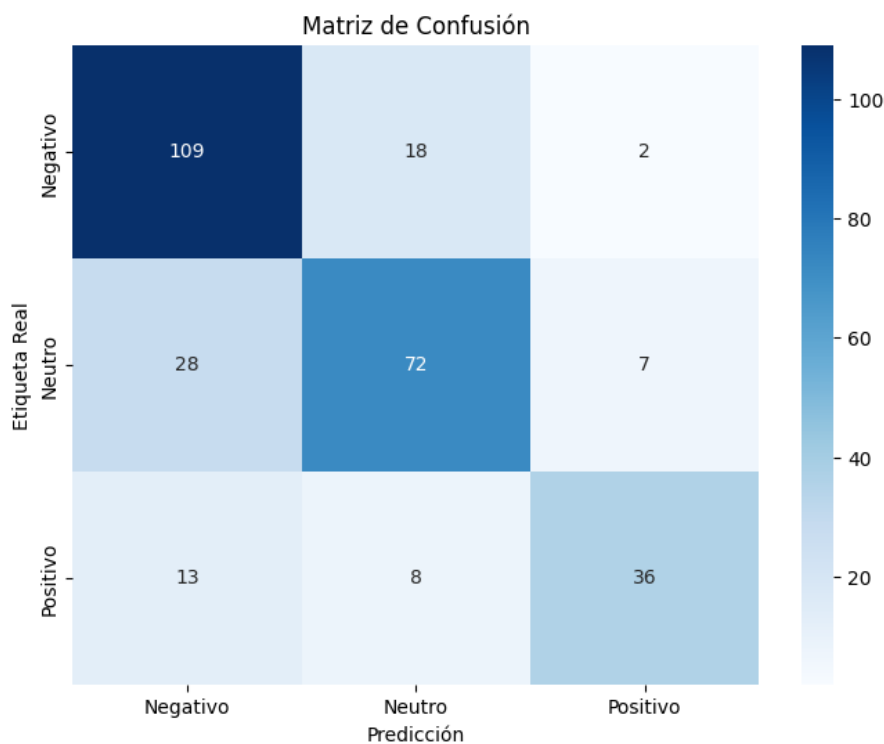
El proceso de entrenamiento y evaluación fue gestionado mediante el objeto Trainer de transformers, que automatizó el manejo de los ciclos de entrenamiento y validación, facilitando el ajuste de los hiperparámetros del modelo. Se probaron múltiples configuraciones para la tasa de aprendizaje, el número de épocas, y el tamaño de batch, lo que contribuyó a afinar la precisión de la clasificación de sentimientos. Finalmente, se establecieron 3 épocas de entrenamiento, un tamaño de batch de 8, una tasa de aprendizaje de 5×10^{-5} , y un weight decay de 0.01. Además, se incluyeron pasos de calentamiento (warmup_steps) y se fijaron otros parámetros, como el directorio de guardado de resultados y el directorio de logs, que ayudaron a controlar el flujo de datos y monitorizar el progreso del entrenamiento. La estrategia de evaluación se definió al final de cada época (evaluation_strategy="epoch"), y también se guardó el modelo al final de cada época (save_strategy="epoch"), asegurando que se cargara el mejor modelo (load_best_model_at_end=True) en función de los resultados de validación.

Para la arquitectura, se empleó un BertConfig especializado, configurado con un tamaño de capa oculta de 768, 12 capas de atención, y 12 capas ocultas, siguiendo el diseño básico de BERT, optimizado para clasificación de secuencias de una sola etiqueta. El BertForSequenceClassification permitió gestionar la salida del modelo en tres etiquetas de sentimiento, conforme a la estructura predefinida en el campo id2label y label2id.

Los resultados obtenidos mostraron una evolución positiva en el rendimiento, con una pérdida de evaluación que se estabilizó en valores bajos a lo largo de las distintas etapas de entrenamiento. Al final del proceso de ajuste, el modelo logró un accuracy del 92,13%, indicando un excelente

desempeño en la clasificación de las frases en las tres categorías de sentimiento: negativo, neutro y positivo. Este alto nivel de precisión refleja la efectividad del enfoque, garantizando que el modelo pudiera generalizar correctamente.

El modelo desarrollado y entrenado se utilizó para evaluar la base de entrenamiento (artículos correspondientes a los años 2020 a 2022), obteniendo un desempeño notable en la clasificación de los sentimientos.

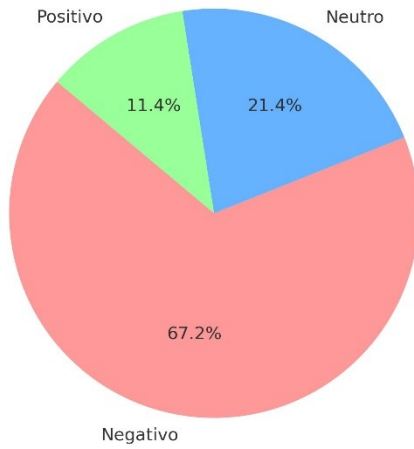


*Ilustración 21 Matriz de confusión Modelo de sentimiento desarrollado
Elaboración por autor*

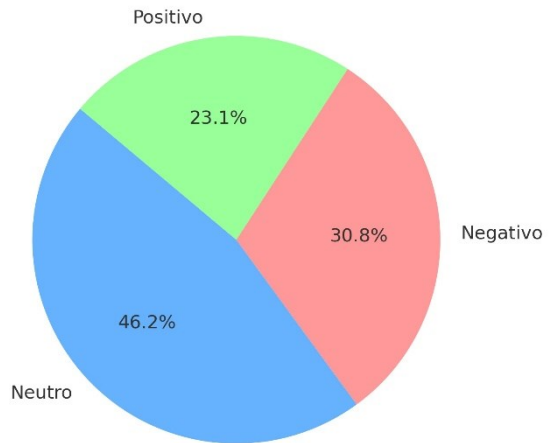
Los resultados del reporte de clasificación mostraron un accuracy global de 0.74, lo que indica que el 74% de las predicciones fueron correctas. En términos de precisión, el modelo presentó un desempeño más alto en la categoría Negativa, con un f1-score de 0.78, reflejando una alta eficacia en la identificación de sentimientos positivos en los artículos. Las categorías Positiva y Neutra alcanzaron f1-scores de 0.71 y 0.70, respectivamente, indicando un buen rendimiento en la clasificación de esos sentimientos, aunque con ligeras variaciones en el recall. En general, los resultados reflejan un buen equilibrio entre las diferentes métricas de evaluación, con una ligera tendencia a detectar más adecuadamente los sentimientos negativos y positivos en comparación con los neutros. Estos resultados validan la efectividad del modelo para analizar el sentimiento en los artículos de prensa en medios digitales.

Una vez comprobada la capacidad del modelo en la clasificación de los artículos correspondientes a los años 2020 a 2022, se procedió a evaluar los artículos del año 2023, que constituyen el objeto principal de esta investigación.

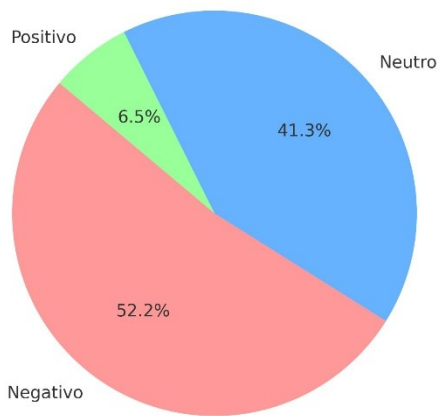
Distribución de Sentimientos El Espectador



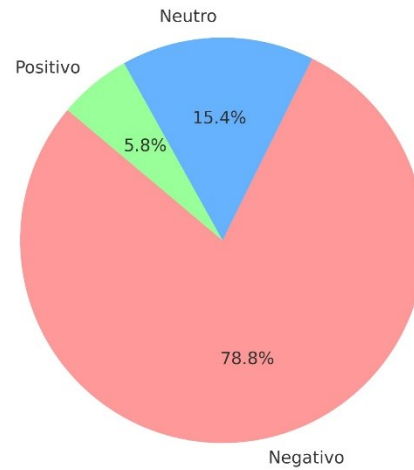
Distribución de Sentimientos La República



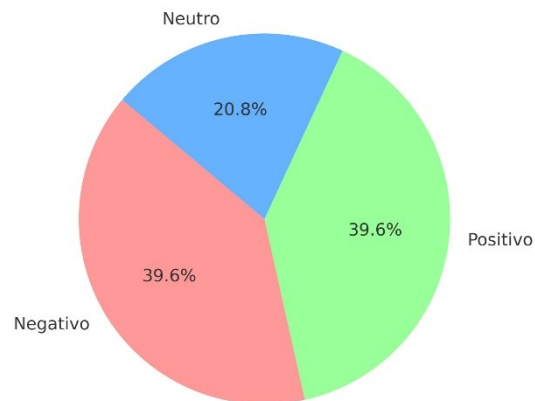
Distribución de Sentimientos Portafolio



Distribución de Sentimientos Semana



Distribución de Sentimientos Las 2 Orillas



*Ilustración 22 Distribución de sentimientos por periódico
Elaborado por autor*

Para un análisis profundo de la percepción de paz en estos periódicos, es crucial tener en cuenta no solo su línea editorial, sino también las condiciones sociopolíticas de Bogotá. La ciudad, como epicentro económico y cultural, es testigo directo de los efectos del conflicto armado, las reformas políticas, las tensiones económicas y la demanda de justicia. Estos factores han alimentado un clima de escepticismo y división sobre la viabilidad de una paz sostenible.

El Espectador con una alta proporción de contenido negativo (67.1%) refleja una postura crítica frente a las promesas de paz y las dificultades en su implementación. Este medio, conocido por su tradición de defensa de derechos humanos y democracia, ha sido un observador de las falencias del Estado en aspectos como justicia y derechos civiles. En un contexto donde las reformas y acuerdos de paz avanzan lentamente, El Espectador enfatiza las deficiencias estructurales en seguridad y gobernanza, lo cual se alinea con la frustración de gran parte de la sociedad bogotana, que percibe que la paz aún es débil y que las instituciones no protegen efectivamente a los ciudadanos.

La República tiene un enfoque económico, y su cobertura refleja la preocupación de la sociedad bogotana y empresarial respecto a cómo la paz impacta la estabilidad económica. Aunque se observa un tono neutro (46.2%) y una considerable presencia de contenido positivo (23.1%), el 30.8% de negatividad señala las barreras que enfrenta la inversión y el crecimiento económico debido a la inseguridad persistente. En Bogotá, donde la economía es un tema prioritario, la percepción de paz entre los sectores empresariales depende de factores como la seguridad jurídica y la eficacia del sistema de justicia, aspectos en los que La República observa que el país aún tiene debilidades que impactan las decisiones económicas.

Las 2 orillas con una proporción equilibrada entre sentimientos positivos y negativos (39.6%) refleja el sentir diverso de la sociedad colombiana, especialmente en Bogotá, una ciudad con una gran variedad de perspectivas ideológicas y sociales. Este medio, que da voz a sectores alternativos y comunitarios, subraya tanto los avances como los desafíos de la paz. En el contexto bogotano, donde existe una gran diversidad de opiniones sobre el proceso de paz, el equilibrio de Las 2 Orillas muestra una narrativa de paz que reconoce tanto los logros en democracia y libertad como los obstáculos, como la desigualdad y la persistencia de grupos armados en las zonas periféricas, que impiden una paz duradera.

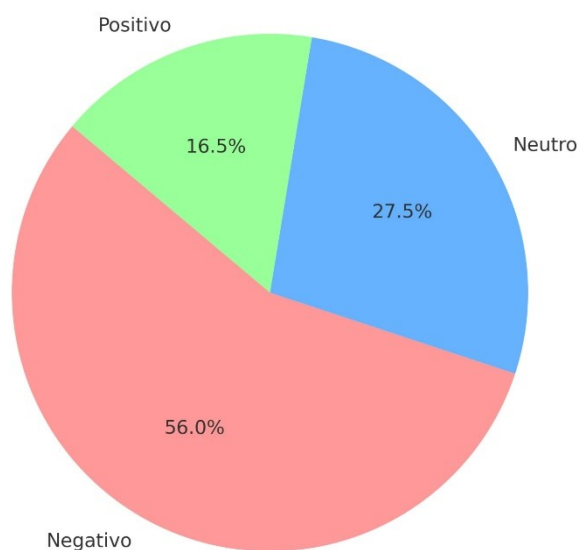
Portafolio, con un 52.2% de contenido negativo y un 41.3% neutro, indica una percepción de paz que está en conflicto con el ideal de seguridad económica y estabilidad. Portafolio, al enfocarse en la economía, sugiere que, mientras la paz no sea palpable en términos de desarrollo y seguridad para el sector privado, la percepción seguirá siendo mayormente negativa. Bogotá, como centro financiero, depende de una paz que permita la inversión y el desarrollo de infraestructuras de manera sostenida. La escasa proporción de sentimiento positivo (6.5%) indica que los actores económicos no perciben aún una paz transformadora y estructural, sino una paz frágil, vulnerable a la incertidumbre política y a la ineficacia en políticas de seguridad.

Semana, con un marcado 78.8% de contenido negativo, refleja una visión escéptica y hasta crítica de los logros de paz en el país. Este medio se ha posicionado históricamente como un observador

cercano del conflicto armado y la política colombiana, y su postura parece alinearse con una visión conservadora y crítica frente a las reformas políticas recientes y los avances en justicia y democracia. En Bogotá, una sociedad donde existen sentimientos contradictorios y sectores que no confían plenamente en el proceso de paz, Semana refuerza una narrativa que enfatiza los problemas estructurales y los riesgos a la libertad y la seguridad. La percepción negativa sugiere un escepticismo sobre si la paz realmente garantiza una democracia funcional y una justicia efectiva en todos los niveles de la sociedad.

Para ilustrar la distribución total de sentimientos en torno a la paz entre los bogotanos, se realizó un análisis unificado de los datos de todos los periódicos en 2023 (ilustración 22). Este análisis consolidado muestra que la mayoría de los artículos tienden hacia un sentimiento negativo: un 56.1% de los artículos se clasifican como negativos, un 27.5% como neutros, y solo un 16.5% como positivos, representando 143, 70 y 42 artículos, respectivamente. A pesar de la diversidad de enfoques en cada medio, esta visión general indica que la cobertura sobre la paz en 2023 tiende a inclinarse hacia un sentimiento negativo, proporcionando un contexto valioso para comprender la percepción de la paz en Bogotá.

Distribución Total de Sentimientos (5 Periódicos Combinados)



*Ilustración 23 Análisis de sentimiento total
Elaboración por autores*

Este sesgo hacia lo negativo en la percepción de paz refleja las complejas realidades socioeconómicas y políticas de Bogotá. La capital enfrenta problemas como la persistente inseguridad, la falta de confianza en el sistema de justicia y la creciente desigualdad, temas que intensifican las preocupaciones sobre la paz y su sostenibilidad. Los artículos negativos, que representan más de la mitad del total, abordan estos desafíos desde una perspectiva crítica, cuestionando la eficacia de las políticas públicas y la falta de avances tangibles en seguridad y justicia para garantizar una paz sostenible.

Por otro lado, el 27.5% de los artículos neutros indica que algunos medios ofrecen un análisis balanceado, discutiendo tanto logros como retos en el proceso de paz sin inclinarse hacia una valoración extrema. Este enfoque neutral aporta matices y responde a la necesidad de una evaluación objetiva de los avances y limitaciones en materia de paz. Finalmente, la cobertura positiva, que abarca solo el 16.5% del total, se centra en logros específicos, tales como avances en derechos humanos, iniciativas de reconciliación y programas de desarrollo, aunque es insuficiente para contrarrestar la percepción negativa dominante.

Esta representación general de la percepción de paz en Bogotá se caracteriza por un decepción y desconfianza generalizados hacia los procesos actuales. La elevada proporción de artículos negativos indica que la ciudadanía vive en un contexto de inseguridad y ansiedad, donde las esperanzas de alcanzar una paz duradera se ven opacadas por las realidades diarias de violencia y desigualdad. Este entorno de incertidumbre se traduce en una crítica a la efectividad de las políticas y acciones gubernamentales, reflejando un deseo colectivo de soluciones concretas que aborden tanto la justicia como la estabilidad. Por lo tanto, la percepción de paz se ve influenciada no solo por los acontecimientos actuales, sino también por la historia reciente del país y las experiencias vividas por los ciudadanos.

4.3 OBJETIVO 3

Crear un indicador con los resultados del modelo desarrollado, en el objetivo específico número 2, que refleje cuantitativamente la percepción de paz derivada de la categorización de expresiones relacionadas al tema en las publicaciones digitales en Bogotá.

4.3.1 Diseño y Desarrollo de un Indicador

La idea central detrás de este indicador es proveer una herramienta numérica sencilla y accesible que refleje, de manera cuantitativa, el balance entre las percepciones positivas, negativas y neutras relacionadas con la percepción de paz. Este indicador ha sido diseñado con un enfoque intuitivo, para garantizar que cualquier lector, independientemente de su nivel de conocimiento técnico, pueda interpretar fácilmente los resultados. Para ello, el indicador se construye en una escala de -1 a 1, donde los valores extremos reflejan posturas claras de rechazo o apoyo a la paz, y el valor intermedio refleja una posición neutral, sin inclinación clara.

La elección de este rango facilita una comprensión clara y directa de la percepción social sobre la paz, lo que permite su interpretación por parte de diversos actores sociales y políticos. Además, este rango resulta útil para analizar la evolución de la percepción de paz a lo largo del tiempo, ya que permite identificar de forma inmediata tendencias hacia una mayor aceptación de los procesos de paz o, en contraste, un incremento en la polarización y el rechazo. Para calcular el indicador, se han desarrollado diversas fórmulas matemáticas que permiten ajustar la medición según las características específicas de los datos analizados. A continuación, se presentan las fórmulas utilizadas para determinar el valor del indicador en cada uno de los modelos propuestos:

Modelo 1: En este modelo, se asignan valores fijos a las categorías de publicaciones: +1 para las

expresiones positivas, 0 para las neutras y -1 para las negativas. El índice de percepción de paz se calcula como el promedio ponderado de estas tres categorías, considerando la proporción de cada una respecto al total de publicaciones. La fórmula es sencilla y directa

$$\text{Índice de Percepción de Paz} = \frac{(N_{\text{positivos}} \times 1) + (N_{\text{neutro}} \times 0) + (N_{\text{negativos}} \times -1)}{N_{\text{total}}}$$

Pros: Este modelo es extremadamente sencillo de entender y calcular, lo que facilita su interpretación, incluso para audiencias no técnicas. Además, la clasificación clara y directa de las publicaciones en tres categorías hace que el proceso sea intuitivo. Este modelo también es rápido de aplicar, lo cual es útil en análisis preliminares o estudios con un gran volumen de datos.

Contras: A pesar de su simplicidad, este modelo tiene limitaciones significativas. Al asignar un valor fijo a cada categoría, no refleja adecuadamente la intensidad de las opiniones. Por ejemplo, una publicación muy positiva o muy negativa tiene el mismo peso que una publicación ligeramente positiva o negativa. Además, las publicaciones neutras no tienen impacto en el resultado, lo que podría ser un inconveniente en contextos donde las posturas neutras también tienen relevancia.

Modelo 2 Este modelo calcula la diferencia entre el número de publicaciones positivas y negativas, normalizada por el total de publicaciones. A esta diferencia se le aplica un ajuste en función de la proporción de publicaciones neutras, con el fin de moderar su impacto en el cálculo.

$$\text{Índice de Percepción de Paz} = \left(\frac{N_{\text{positivos}} - N_{\text{negativos}}}{N_{\text{total}}} \right) \times \left(1 - \frac{N_{\text{neutro}}}{N_{\text{total}}} \right)$$

Pros: Este modelo tiene la ventaja de considerar tanto las publicaciones positivas como negativas, lo que permite un análisis más equilibrado de las percepciones de paz. Al ponderar la proporción de publicaciones neutras, también se ajusta la influencia de aquellas publicaciones que no expresan una postura clara. Esta característica le da al modelo una mayor flexibilidad y precisión en contextos con opiniones divididas.

Contras: Sin embargo, el modelo es más sensible a las variaciones en las publicaciones neutras, lo que podría hacer que el indicador sea más inestable en contextos donde las opiniones neutras sean prevalentes. Además, aunque ajusta el impacto de las neutras, no ofrece una consideración

profunda de la intensidad de las expresiones positivas o negativas, lo que podría ser un punto débil en contextos más complejos.

Modelo 3 En este modelo, se calculan las probabilidades de expresiones positivas, negativas y neutras, basadas en su proporción respecto al total de publicaciones. Luego, el índice se ajusta para moderar la influencia de las publicaciones neutras, ponderándolas de forma más suave

$$P_{positivo} = \frac{N_{positivo}}{N_{total}} \quad P_{negativo} = \frac{N_{negativo}}{N_{total}} \quad P_{neutro} = \frac{N_{neutro}}{N_{total}}$$

$$Indice\ de\ Percepción\ de\ Paz = P_{positivo} - P_{negativo} + \frac{P_{neutro}}{2}$$

Pros: Este modelo ofrece un enfoque más matizado, al moderar el impacto de las publicaciones neutras. Esto resulta útil para evitar que un gran volumen de publicaciones neutras distorsione de manera desproporcionada la percepción general. Además, al trabajar con probabilidades en lugar de simples conteos, el modelo es más robusto frente a variaciones en el número total de publicaciones, lo que lo hace más confiable en análisis de gran escala.

Contras: Sin embargo, este modelo es ligeramente más complejo que el Modelo 1, lo que podría dificultar su interpretación para audiencias no especializadas. Además, el hecho de que las publicaciones neutras se ponderen a la mitad de su valor podría no reflejar de manera justa su impacto real en el contexto sociopolítico de la ciudad, especialmente si se considera que las opiniones neutras pueden tener un papel importante en la dinámica social.

Modelo 4 El cuarto modelo también calcula la diferencia entre las publicaciones positivas y negativas, pero a diferencia del Modelo 2, ajusta de manera no lineal el impacto de las publicaciones neutras mediante una fórmula de normalización.

$$P_{positivo} = \frac{N_{positivo}}{N_{total}} \quad P_{negativo} = \frac{N_{negativo}}{N_{total}} \quad P_{neutro} = \frac{N_{neutro}}{N_{total}}$$

$$Indice\ de\ Percepción\ de\ Paz = \frac{P_{positivo} - P_{negativo}}{1 + P_{neutro}}$$

Pros: Este modelo tiene la ventaja de evitar que un alto volumen de publicaciones neutras distorsione la percepción de paz, lo que lo hace más robusto y estable. La normalización de las neutras también asegura que su influencia no sea excesiva, lo que resulta útil en contextos donde las opiniones claras son más relevantes para la evaluación de la paz. Este enfoque es especialmente valioso en escenarios de polarización, donde la distorsión causada por las neutras puede ser un problema.

Contras: Aunque proporciona un ajuste más preciso, este modelo es más matemáticamente complejo que los anteriores, lo que puede dificultar su comprensión y uso por parte de un público general. Además, el modelo podría subestimar la importancia de las publicaciones neutras en ciertos contextos, donde estas puedan jugar un papel crucial en la mediación de las posturas extremas.

Con el fin de ponderar adecuadamente la influencia de cada medio en el cálculo del Índice de Percepción de Paz, se realizó una normalización lineal inversa basada en la popularidad de los periódicos, utilizando el ranking proporcionado por SCImago Lab. Este ajuste tiene como objetivo equilibrar el impacto de los diferentes periódicos, considerando que aquellos con mayor popularidad reflejan la percepción de una mayor cantidad de personas, mientras que los de menor popularidad tienen un impacto reducido, aunque no nulo. La normalización ajusta los valores del índice en un rango entre 0 y 1. En este contexto, el periódico con mayor popularidad (es decir, el que ocupa una posición más alta en el ranking) recibe un valor de 1, mientras que el periódico con menor popularidad (el que ocupa la posición más baja en el ranking) recibe un valor de 0. Los demás periódicos se asignan valores intermedios según su posición relativa en el ranking. Para evitar que el periódico con la menor popularidad obtenga un valor de 0 absoluto, se consideró la inclusión de un "periódico adicional" en la última posición del ranking, ajustando así el cálculo para que este periódico aún conserve un valor mayor que cero, aunque reducido.

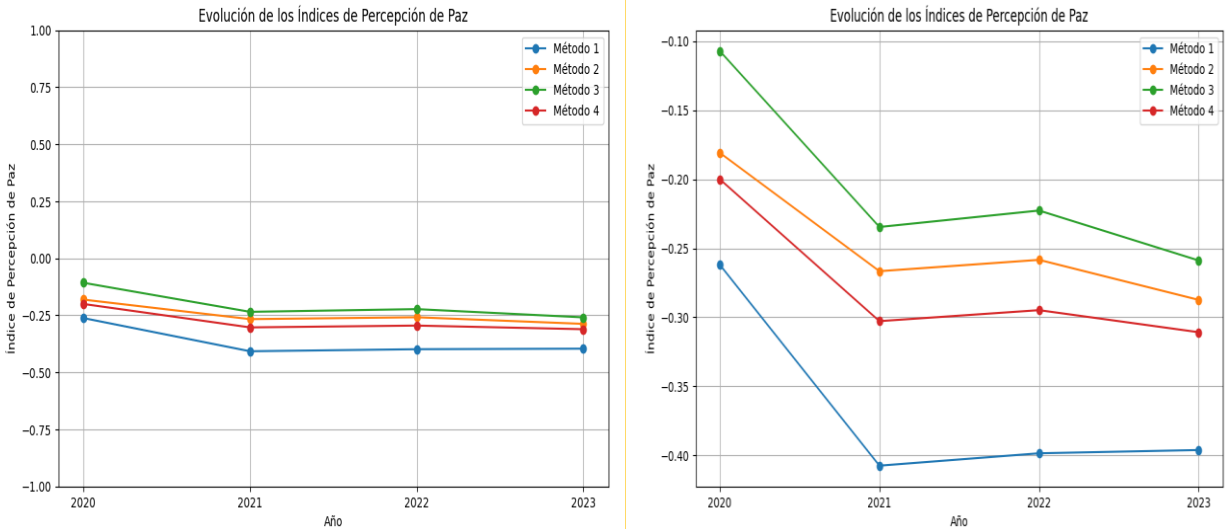
La fórmula empleada para la normalización fue la siguiente:

$$Popularidad\ normalizada = 1 - \frac{Ranking - 1}{Max\ Ranking - 1}$$

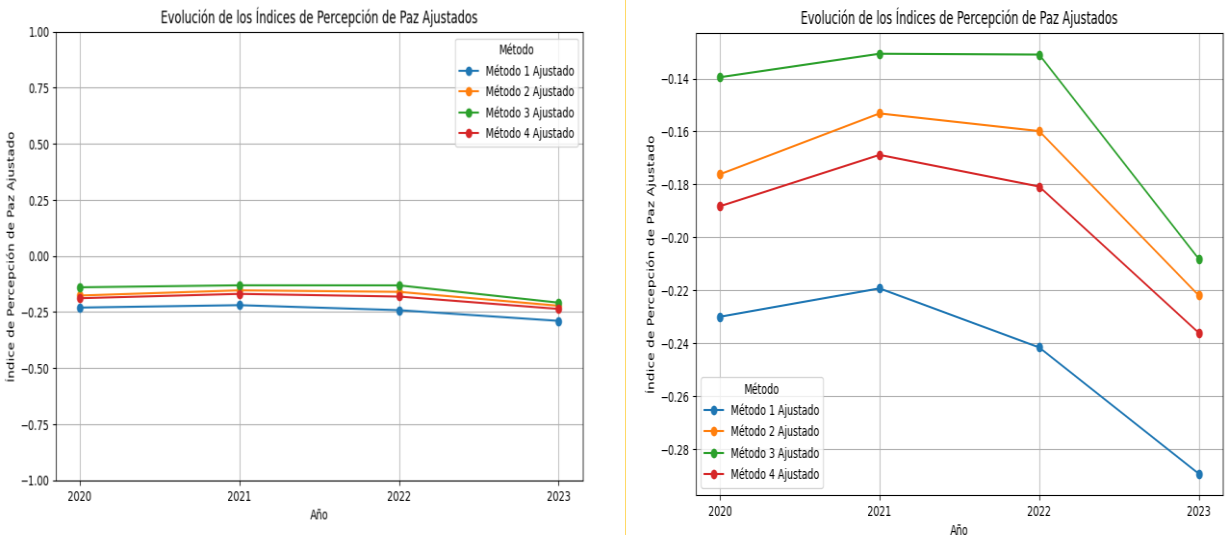
Este proceso de normalización permite ponderar los índices de percepción de paz de cada periódico de acuerdo con su influencia relativa. Posteriormente, cada uno de los índices de percepción de paz propuestos en los modelos fue multiplicado por el valor de popularidad normalizada de su respectivo periódico. Esto ajustó el impacto de cada periódico en el cálculo global, logrando un ajuste proporcional según la influencia de cada medio en la opinión pública. Finalmente, para interpretar los resultados, se calculó un promedio ajustado anual para cada índice de percepción de paz, lo que permitió evaluar la percepción de paz en todos los periódicos

de un mismo año de manera equilibrada. Este ajuste proporcional asegura que los medios con mayor popularidad, que reflejan la percepción de un mayor número de personas, tengan una mayor representación en el índice general, mientras que los de menor popularidad continúan contribuyendo al análisis, pero con un impacto proporcionalmente reducido.

INDICADORES ANTES DE NORMALIZAR



INDICADORES DESPUES DE NORMALIZAR



*Ilustración 24 Comparación de los modelos propuestos como índice de percepción de paz
Elaboración por autores*

Al analizar las gráficas generadas por los cuatro modelos, se observa que todos reflejan un comportamiento similar en la percepción de paz en Bogotá. Aunque las diferencias entre ellos radican principalmente en la forma en que ponderan las publicaciones neutrales y ajustan la relación entre las publicaciones positivas y negativas, las tendencias generales y los cambios en la percepción se mantienen consistentes. Este hallazgo sugiere que, independientemente del

enfoque específico utilizado, los modelos capturan de manera adecuada las dinámicas principales de la percepción de paz expresadas en las publicaciones analizadas. Sin embargo, El Modelo 2 fue seleccionado para el cálculo del Índice de Percepción de Paz en Bogotá debido a su capacidad para abordar de manera efectiva los distintos componentes emocionales de las publicaciones. Las razones para su elección incluyen los siguientes factores clave:

1. Equilibrio entre publicaciones positivas y negativas: El Modelo 2 calcula la diferencia entre las publicaciones positivas y negativas, lo cual es fundamental para medir la percepción de paz de manera efectiva. La paz es un concepto cargado emocionalmente, y el contraste entre los sentimientos positivos y negativos es clave para identificar si la población en Bogotá está experimentando una tendencia hacia un ambiente pacífico o, por el contrario, hacia uno de tensión o conflicto. Este modelo establece un equilibrio directo entre ambas categorías, lo que permite una medición clara y comprensible.
2. Incorporación de las publicaciones neutrales: El Modelo 2 incluye un ajuste específico para las publicaciones neutrales, restando su proporción del total de publicaciones. Esto es importante porque las publicaciones neutras, si bien no aportan directamente a la percepción positiva o negativa, pueden diluir o suavizar los efectos de las publicaciones emocionales. Al restar su influencia, el modelo refleja de manera más fiel el sentimiento real de la población en relación con la paz.
3. Simplicidad y efectividad: El modelo es relativamente sencillo, lo que facilita su interpretación. Al combinar la diferencia de publicaciones positivas y negativas con la proporción de publicaciones neutrales, el modelo ofrece una medición clara y fácilmente comprensible, sin complejidades adicionales que podrían generar dificultades en la interpretación de los resultados.
4. Ajuste de las publicaciones según su relevancia: Al calcular el índice tomando en cuenta la proporción de publicaciones neutrales, el Modelo 2 refleja una percepción más precisa de la paz, ya que las publicaciones neutras a menudo representan una falta de opinión fuerte sobre el tema y no contribuyen directamente a la percepción de paz. Este ajuste permite que el índice se enfoque más en las publicaciones que realmente expresan una opinión clara sobre la paz, ya sea positiva o negativa.
5. Normalización por la popularidad de los medios: Además, este modelo permite incorporar la normalización lineal inversa basada en la popularidad de los periódicos, lo que ajusta el impacto de cada medio según su relevancia en la opinión pública. Al ponderar las publicaciones de los medios más populares con un mayor peso, se asegura que el índice refleje la percepción de paz de un mayor número de personas, evitando que los medios menos populares distorsionen el resultado final.

En conjunto, el Modelo 2 proporciona una herramienta eficaz para medir la percepción de paz en Bogotá, siendo capaz de reflejar la interacción entre las publicaciones positivas, negativas y

neutrales, y ajustando la influencia de los medios de acuerdo con su popularidad. A continuación, se presenta una descripción detallada de los componentes y características que componen este indicador.

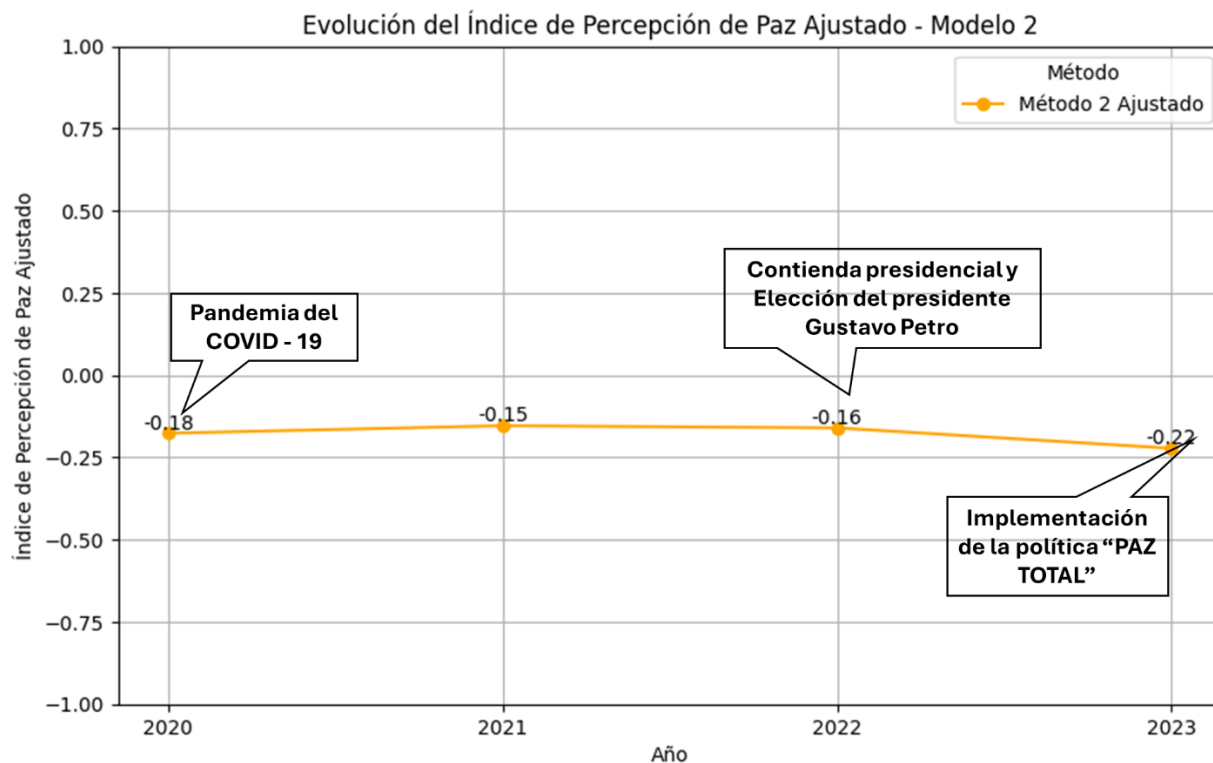
Hoja de vida del indicador	
Nombre	Índice de Percepción de Paz en Bogotá
Descripción	Este indicador mide la percepción de paz en Bogotá a través del análisis de sentimientos en las publicaciones digitales de los principales medios de comunicación. Se calcula considerando las publicaciones positivas y negativas, mientras que las publicaciones neutras se restan proporcionalmente para obtener una evaluación más precisa del ambiente de paz. Además, se ajusta por la popularidad de los periódicos para reflejar la influencia de los medios más populares en la percepción pública.
Unidad de medida	Adimensional (rango entre -1 y 1)
Formula	$\text{Indice percepcion de paz} = \frac{N_{positivos} - N_{negativo}}{N_{total}} \times \left(1 - \frac{N_{neutro}}{N_{total}}\right) \times \left(1 - \frac{\text{Ranking} - 1}{\text{Max Ranking} - 1}\right)$
Variables del indicador	<p>Npositivos: Número de publicaciones con expresión positiva respecto a la paz. Nnegativos: Número de publicaciones con expresión negativa respecto a la paz. Nneutro: Número de publicaciones con expresión neutral sobre paz. Ntotal: Número total de publicaciones analizadas. Ranking: Posición de popularidad del periódico en el ranking. Max Ranking: Posición del periódico menos popular en el ranking.</p>
Periodicidad de la medición del indicador	Anual. El indicador se calcula al final de cada año, utilizando las publicaciones digitales correspondientes a ese período.
Tipología	Indicador de resultado
Fecha de creación	Noviembre 2024

*Tabla 4: Ficha técnica del indicador desarrollado
Elaborada por autores*

4.3.2 Evaluación y Análisis de los Resultados Obtenidos con el Indicador Diseñado

El desarrollo del Índice de Percepción de Paz en Bogotá busca proporcionar una herramienta que permita capturar la percepción pública sobre la paz en la capital colombiana. Este indicador

combina técnicas de análisis de sentimientos aplicadas a publicaciones digitales de medios de comunicación, ponderadas por factores como la popularidad de las fuentes, para generar una representación equilibrada y matizada de esta percepción. Su análisis no solo refleja la respuesta ciudadana a eventos y políticas relacionadas con la paz, sino que también resalta cómo los discursos mediáticos y los contextos locales reflejan las percepciones colectivas. A continuación, se presenta una evaluación detallada de los resultados obtenidos con este indicador, integrando además los datos de seguridad ciudadana para enriquecer el entendimiento de su comportamiento durante el periodo 2020-2023.



*Ilustración 25: Evolución del índice de percepción de paz desarrollado
Elaboración por autores*

A lo largo de los cuatro años analizados, el índice de percepción de paz en Bogotá se mantiene dentro de un **rango negativo moderado**, pero cerca del valor neutro (cero). Esto sugiere que, aunque los ciudadanos experimentan algunos problemas de seguridad y confianza en las instituciones, la situación no ha sido lo suficientemente grave como para generar una percepción de paz totalmente deteriorada. En general, el comportamiento del indicador desarrollado revela una notable estabilidad a lo largo de los años, lo que sugiere que los factores subyacentes que afectan la percepción de paz en Bogotá no han experimentado grandes variaciones. Esta estabilidad en las mediciones podría indicar que el contexto sociopolítico y la situación de seguridad en la ciudad, aunque en constante evolución, no han experimentado cambios abruptos que alteren significativamente la percepción colectiva sobre la paz. Sin embargo, los pequeños cambios en las tendencias observadas entre los años, aunque leves, pueden tener un significado más profundo. Estos cambios reflejan las fluctuaciones más sutiles en la opinión pública, lo que

está relacionado con acontecimientos específicos que marcan la agenda pública y las percepciones de los ciudadanos. A continuación, se expresarán las condiciones y los principales eventos sociopolíticos de Bogotá durante los años estudiados que impactaron la percepción de paz.

La pandemia del COVID-19

La pandemia del COVID-19 marcó un antes y un después en la vida de los colombianos, afectando todos los aspectos de la sociedad, desde la salud hasta la economía y la seguridad. En términos de percepción de paz, el análisis de los resultados del indicador propuesto refleja cómo este evento global influyó negativamente en la sensación de bienestar y estabilidad de los ciudadanos. En los primeros años de la pandemia (2020-2021), se observa una ligera mejora en la percepción de paz, con un índice que se mantiene cerca de los valores negativos moderados. Esto podría interpretarse como una respuesta inicial a la sensación de incertidumbre y de cierre de actividades, lo que llevó a la población a experimentar una reducción en las tensiones sociales. Sin embargo, la estabilidad que alcanzó el indicador fue efímera.

A medida que la pandemia se extendió y los efectos económicos y sociales se hicieron más evidentes, la percepción de paz comenzó a deteriorarse de nuevo. En 2022 y 2023, el índice de percepción mostró un descenso, alcanzando un valor de -0.22 en 2023. Este descenso puede estar relacionado con las dificultades económicas derivadas de la crisis sanitaria, las crecientes tensiones políticas y sociales, y el miedo generalizado a la propagación del virus. Aunque la implementación de medidas de confinamiento y restricciones sociales inicialmente redujo la actividad criminal en algunas zonas, la crisis económica y la creciente desigualdad generaron un ambiente de incertidumbre y malestar que afectó la percepción general de seguridad.

Es importante destacar que, a pesar de estas fluctuaciones, la percepción de paz en Colombia nunca alcanzó valores extremadamente negativos, lo que podría sugerir que la sociedad colombiana, aunque afectada, logró mantener cierta resiliencia frente a los desafíos presentados por la pandemia. En lugar de caer en una crisis total de seguridad o desconfianza, los ciudadanos parecieron adaptarse a las nuevas realidades, reconociendo las dificultades, pero manteniendo un cierto nivel de esperanza y estabilidad. Sin embargo, el impacto del COVID-19 dejó en evidencia que los problemas estructurales en Colombia, como la desigualdad y la violencia, siguen siendo factores determinantes en la percepción de paz, y aunque la situación mejoró en algunos aspectos, las cicatrices dejadas por la pandemia continúan siendo un desafío para la construcción de una paz duradera.

La elección del presidente Gustavo Petro y su política de trabajo “Paz Total”

Las contiendas presidenciales previas a la elección de Gustavo Petro estuvieron marcadas por una profunda polarización política y social en Colombia. Durante el periodo de campaña, temas como la implementación de los acuerdos de paz con las FARC, la seguridad nacional, y la desigualdad económica ocuparon un lugar central en el debate público. Los candidatos representaron posturas contrastantes: desde quienes defendían una línea continuista y de mano dura frente a los grupos armados, hasta propuestas más progresistas, como las de Petro, que buscaban priorizar el diálogo y la transformación social. Esta polarización no solo evidenció las tensiones latentes en la sociedad colombiana, sino que también creó un ambiente de incertidumbre que influyó en la percepción de paz en el país. Mientras unos sectores veían en el cambio político una oportunidad para construir un futuro más pacífico y equitativo, otros temían que un giro hacia la izquierda pudiera poner en riesgo la estabilidad institucional y exacerbar los problemas de seguridad. Estas tensiones, sumadas a la intensificación del discurso político, dejaron un impacto en la percepción de paz que se reflejó en los índices de los años previos a la elección presidencial.

La elección de Gustavo Petro como presidente de Colombia en 2022 marcó un hito en la política colombiana, ya que fue el primer presidente de izquierda en la historia del país. Su campaña electoral estuvo centrada en una propuesta ambiciosa y transformadora: la política de "Paz Total", que prometía avanzar en la implementación de acuerdos de paz con diversos grupos armados, así como en la construcción de una sociedad más equitativa y menos violenta. Sin embargo, la implementación de esta política, al igual que las expectativas generadas en la campaña, ha tenido un impacto considerable en la percepción de paz en Colombia, especialmente a partir de 2022 y durante los primeros años de su mandato.

A pesar de la esperanza que generó la propuesta de "Paz Total", los resultados obtenidos han sido mixtos y, en algunos casos, frustrantes para la población. El índice de percepción de paz, reflejado en el indicador propuesto, mostró una tendencia moderadamente negativa en 2022 (-0.16) y 2023 (-0.22), lo que sugiere que, a pesar de las intenciones del gobierno, la situación de seguridad y el clima de violencia no mejoraron de manera significativa en el corto plazo. De hecho, algunos de los intentos de negociación con grupos armados y disidencias generaron reacciones mixtas entre la ciudadanía, que, por un lado, anhelaba la paz, pero, por otro, temía que estas negociaciones pudieran ser vistas como una concesión a actores violentos.

El fracaso parcial de la política de "Paz Total" también puede reflejarse en la frustración de muchos colombianos que, a pesar de la promesa de cambio, no han visto mejoras sustanciales en la seguridad en sus territorios. En lugar de una disminución de la violencia, algunos sectores del país han experimentado un aumento de las confrontaciones armadas y de los desplazamientos forzados. Este panorama ha afectado negativamente la percepción de paz, como lo indican los resultados de los índices. La expectativa de un proceso de paz integral se ha visto opacada por la

realidad de los desafíos estructurales, como el narcotráfico, la pobreza, y la falta de confianza en las instituciones, que siguen siendo barreras importantes para lograr una paz duradera.

El aumento de la violencia en algunas regiones y la percepción de que las negociaciones podrían haber favorecido a grupos armados generaron desconfianza en la población, especialmente en aquellos que temen que se estén tomando decisiones que podrían poner en riesgo la seguridad de las comunidades más vulnerables. Este contexto ha hecho que la percepción de paz en Colombia se mantenga en niveles moderadamente negativos, reflejando una sensación de desconfianza y falta de progreso en el proceso de pacificación que la campaña electoral de Petro había prometido.

En resumen, aunque la política de "Paz Total" tenía un objetivo noble y transformador, la implementación de esta ha estado marcada por dificultades y desafíos. Las expectativas generadas durante la campaña electoral no se han traducido en mejoras inmediatas en la seguridad, y los efectos de la política no han logrado cambiar de manera significativa la percepción de paz en la sociedad colombiana. La percepción se mantiene en niveles negativos, con una ciudadanía que, si bien aún espera el cumplimiento de las promesas de paz, se enfrenta a la dura realidad de la persistente violencia y la inseguridad en varias regiones del país.

El proceso de paz en Colombia desde 2020 al 2023

La línea del tiempo del proceso de paz en Colombia, comprendida entre 2020 y 2023, presenta los momentos clave que han determinado el progreso y la implementación de los acuerdos de paz, así como los desafíos y obstáculos que han surgido en este proceso, reflejando tanto los avances como las dificultades en su ejecución.

Año	Fecha	Acontecimiento
2020	09/02/2020	La Corte Constitucional de Colombia avala la Jurisdicción Especial para la Paz (JEP), reafirmando su rol en el cumplimiento de los acuerdos de paz.
	05/11/2020	El Instituto Kroc reporta que solo el 28% de lo acordado en el Acuerdo de Paz se ha implementado hasta la fecha.
	03/12/2020	Naciones Unidas y organizaciones de la sociedad civil denuncian el asesinato de 292 excombatientes de las FARC desde la firma del Acuerdo de Paz.

	13/03/2020	La Organización Mundial de la Salud (OMS) declara la pandemia global por COVID-19, afectando el contexto social y político de Colombia.
2021	01/03/2021	Naciones Unidas alerta sobre el incremento en los asesinatos de líderes sociales, alcanzando 152 casos en el año.
	15/06/2021	La Cruz Roja reporta un aumento del 148% en desplazamientos forzados respecto al año anterior.
	09/11/2021	Escalamiento de violencia afecta la reincorporación de excombatientes con más de 46 cuerpos no identificados recuperados en zonas de conflicto.
	19/06/2021	Elecciones presidenciales en Colombia, con alta polarización política. Gustavo Petro se perfila como uno de los principales candidatos en el marco de la Paz Total.
	31/12/2021	Fin de las restricciones más estrictas de la pandemia de COVID-19 en Colombia.
	2022	27/01/2022
07/03/2022		El Pacto Histórico comienza a promover la Paz Total.
14/04/2022		El programa de gobierno de Gustavo Petro habla de la Paz Total.
25/07/2022		Danilo Rueda es nombrado Alto Comisionado para la Paz.
25/08/2022		Ministerio de Defensa anuncia el fin de los bombardeos.
30/08/2022		Se radica la Ley de Paz Total (prórroga de la Ley de Orden Público).
02/10/2022		OACP afirma que 10 grupos armados se han sumado a la Paz Total.
04/11/2022		Sanción presidencial de la Ley 2272 de 2022 "Paz Total".
05/12/2022		Gustavo Petro promete excarcelar a miembros de la Primera Línea para ser gestores de paz.
09/12/2022		Se crea la Comisión Intersectorial para la Promoción de la Paz, la Reconciliación y la Participación Ciudadana.

	31/12/2022	Se crea la Instancia de Alto Nivel compuesta por el ministro de Defensa, el Alto Comisionado para la Paz y el director nacional de Inteligencia.
2023	28/01/2023	Se revela escándalo de cartel de abogados que engaña a personas privadas de la libertad.
	15/02/2023	Gobierno presenta primer borrador de Ley de Sometimiento.
	09/03/2023	Consejo de Política Criminal da concepto favorable a la Ley de Sometimiento.
	15/03/2023	Radicación oficial del proyecto de ley de sometimiento a la justicia.
	10/04/2023	Presidente y fiscal general se reúnen a propósito de la Ley de Sometimiento.
	31/05/2023	Radicación de ponencia de Ley de Sometimiento.
	26/06/2023	Ponencia de magistrado de la Corte Constitucional buscaría declarar inexecutable la Ley de Paz Total.
	14/07/2023	La Corte Constitucional no se pronuncia de fondo sobre declarar inexecutable la Ley de Paz Total.
	13/03/2023	Comienza la aplicación de la Reforma Tributaria impulsada por el gobierno de Gustavo Petro, con énfasis en recursos para la implementación de acuerdos y la promoción de la justicia social.
	19/06/2023	Elecciones presidenciales, Gustavo Petro confirmado como presidente electo, continuando con la Paz Total.

*Tabla 5 Línea de tiempo proceso de paz en Colombia 2020 – 2023
Elaborada por autores*

Entre 2020 y 2021, el indicador muestra un ligero incremento. Este resultado está relacionado con la consolidación de políticas de implementación del Acuerdo de Paz y con el respaldo internacional a los mecanismos de justicia transicional, como la Jurisdicción Especial para la Paz (JEP) [37]. Durante este periodo, aunque persistieron desafíos, las iniciativas dirigidas a la reincorporación de excombatientes y la ampliación de programas para víctimas generaron un leve aumento en las publicaciones positivas. No obstante, la presencia de publicaciones neutrales, asociadas a incertidumbres sobre la sostenibilidad de estas iniciativas, moderó significativamente el impacto del optimismo reflejado en los medios.

En contraste, entre 2021 y 2022 se registra una pequeña reducción del índice. Este comportamiento se explica por un aumento proporcional en las publicaciones negativas, impulsado principalmente por un incremento en los asesinatos de líderes sociales y defensores de derechos humanos, así como por reportes de violencia en territorios rurales [38]. A pesar de ciertos avances, como la mayor visibilidad de los compromisos gubernamentales en foros internacionales, las publicaciones neutras y negativas ganaron peso en el indicador. La polarización política en el marco de las elecciones presidenciales también amplificó la proporción de publicaciones críticas, afectando el balance general del índice.

Finalmente, entre 2022 y 2023 se observa una reducción más notable en el indicador. Este descenso está correlacionado con el cambio en la dinámica política y social tras la implementación de la política de "Paz Total" [39]. Aunque esta estrategia buscaba ampliar el alcance del proceso de paz a otros actores armados, la percepción de su efectividad se polarizó, reflejándose en un aumento de publicaciones negativas. Los altos niveles de violencia, particularmente el asesinato de excombatientes y la persistencia de desplazamientos forzados, dominaron la narrativa mediática, mientras que las publicaciones positivas relacionadas con la firma de nuevos acuerdos preliminares no lograron compensar el pesimismo predominante. Además, la proporción de publicaciones neutras, que capturaron la incertidumbre frente a los resultados de estas políticas, contribuyó a atenuar cualquier impacto positivo.

El cálculo del índice evidencia cómo el ajuste por publicaciones neutras y la ponderación por popularidad de los medios mitigan la influencia de extremos en la percepción pública, proporcionando una medición equilibrada y representativa. Sin embargo, los cambios observados durante el periodo analizado reflejan que, aunque hay avances en la construcción de paz, la percepción pública sigue siendo altamente sensible a los eventos violentos, la incertidumbre política y los desafíos en la implementación de políticas estructurales. Este comportamiento destaca la necesidad de acciones sostenibles y visibles que transformen tanto la realidad de los territorios como la narrativa mediática sobre la paz en Bogotá.

Indicadores de Seguridad Ciudadana en Bogotá

Los indicadores de seguridad ciudadana en Bogotá son herramientas para evaluar la evolución de la seguridad en la ciudad, permitiendo identificar tanto los avances como los desafíos en la lucha contra la delincuencia [40]. Estos indicadores incluyen aspectos clave como homicidios y hurtos, los cuales se analizan junto con el comportamiento del índice de percepción de paz. Al observar cómo se interrelacionan estos datos, se obtiene una visión más completa de la dinámica social y de seguridad en la capital

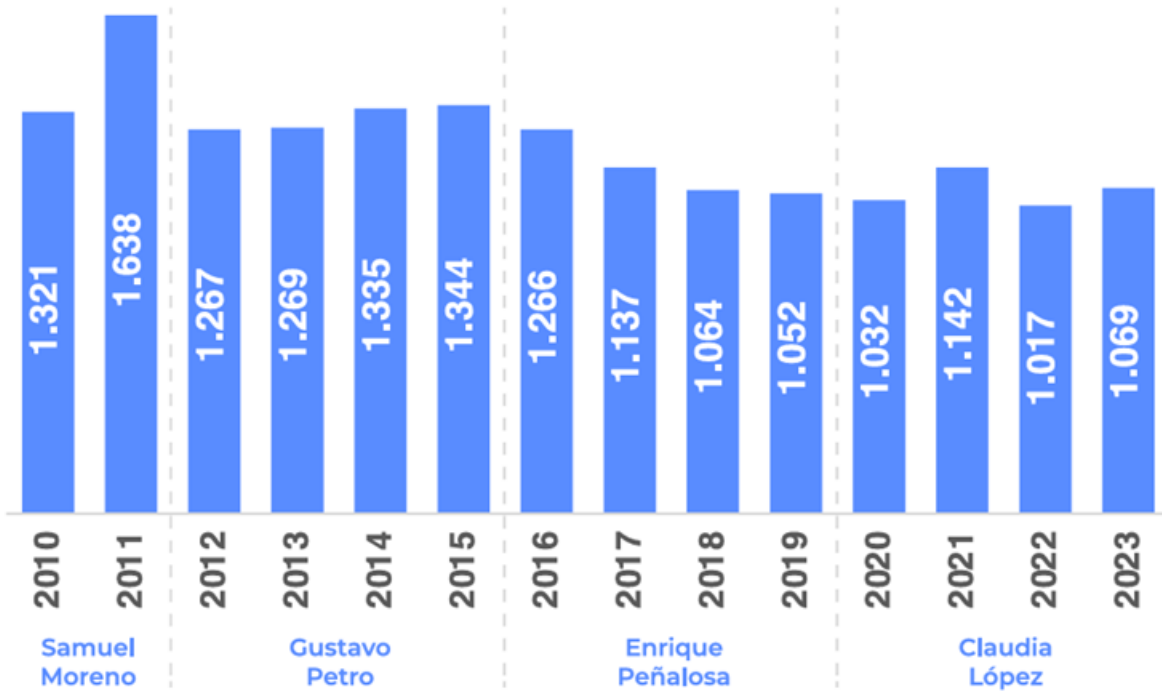


Ilustración 26: Registro histórico de homicidios en Bogotá 2010 -2023

Elaborado por Sistema Estadístico, Delincuencial, Contravencional y Operativo SIEDCO (datos extraídos el 1 febrero de 2024 de la policía nacional)

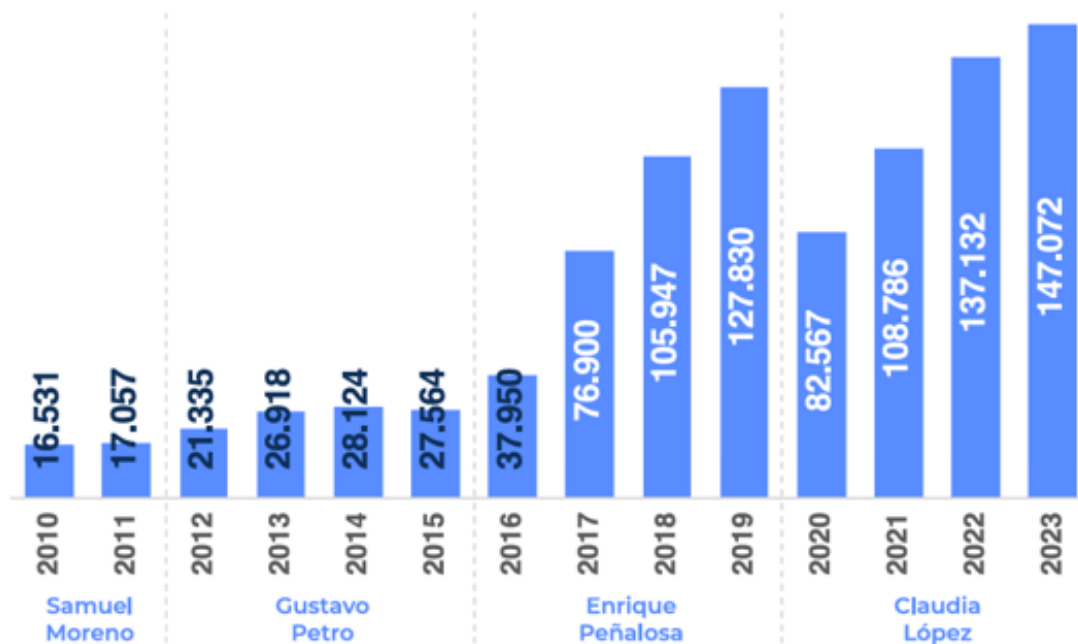


Ilustración 27: Registro histórico de hurtos en Bogotá 2010 – 2023

Elaborado por Sistema Estadístico, Delincuencial, Contravencional y Operativo SIEDCO (datos extraídos el 1 febrero de 2024 de la policía nacional)

Entre 2020 y 2021, los homicidios en Bogotá mostraron un ligero incremento, alcanzando 1,142 casos frente a los 1,030 de 2020, lo que representó un aumento del 10,6%. Este aumento, a pesar de la histórica reducción de homicidios en 2020, se produjo en un contexto de reactivación de actividades sociales y económicas tras el confinamiento. A pesar de este incremento, la percepción de paz mostró una ligera mejora, lo cual podría estar relacionado con las políticas de seguridad implementadas por el gobierno y los esfuerzos de prevención de violencia. No obstante, el repunte de delitos, como el hurto a personas, que creció un 31,75%, podría haber generado una sensación ambigua entre los ciudadanos, donde la percepción de seguridad fue moderada por el aumento de delitos comunes [40].

Entre 2021 y 2022, se registró una pequeña reducción en el índice de percepción de paz, lo cual coincide con una disminución en los homicidios a 1,017 casos, lo que representó una caída del 10,9% respecto al año anterior. Sin embargo, el hurto a personas continuó con una tendencia ascendente, con un aumento del 26,06% en comparación con 2021, alcanzando los 136,209 casos. Este comportamiento refleja una mejora en la violencia letal, pero al mismo tiempo, un incremento en los delitos comunes que pudo haber contribuido a una mayor percepción de inseguridad. Aunque los homicidios descendieron, el aumento en los hurtos en varias localidades de la ciudad afectó la sensación de seguridad, lo que llevó a una ligera disminución en el indicador de paz, particularmente debido a la alta tasa de robos reportados en sectores como Santa Fe y Los Mártires [40].

Finalmente, entre 2022 y 2023, el índice de percepción de paz experimentó una reducción más significativa. Durante este periodo, los homicidios aumentaron nuevamente a 1,069, lo que representó un incremento del 5% respecto al año anterior. Este aumento, combinado con el máximo histórico alcanzado en los hurtos a personas, con 147,072 casos, intensificó la sensación de inseguridad en la ciudad. La violencia, en particular en algunas localidades como La Candelaria y Los Mártires, y el alto número de hurtos reflejaron una mayor preocupación entre la ciudadanía, que se tradujo en un descenso en la percepción de paz [40]. A pesar de los esfuerzos de seguridad, las fluctuaciones en los homicidios y el creciente número de robos contribuyeron a un deterioro más pronunciado en la percepción pública de la seguridad en Bogotá, lo que resalta la necesidad de estrategias más efectivas y diferenciadas para garantizar la seguridad y mejorar la percepción de paz en la ciudad.

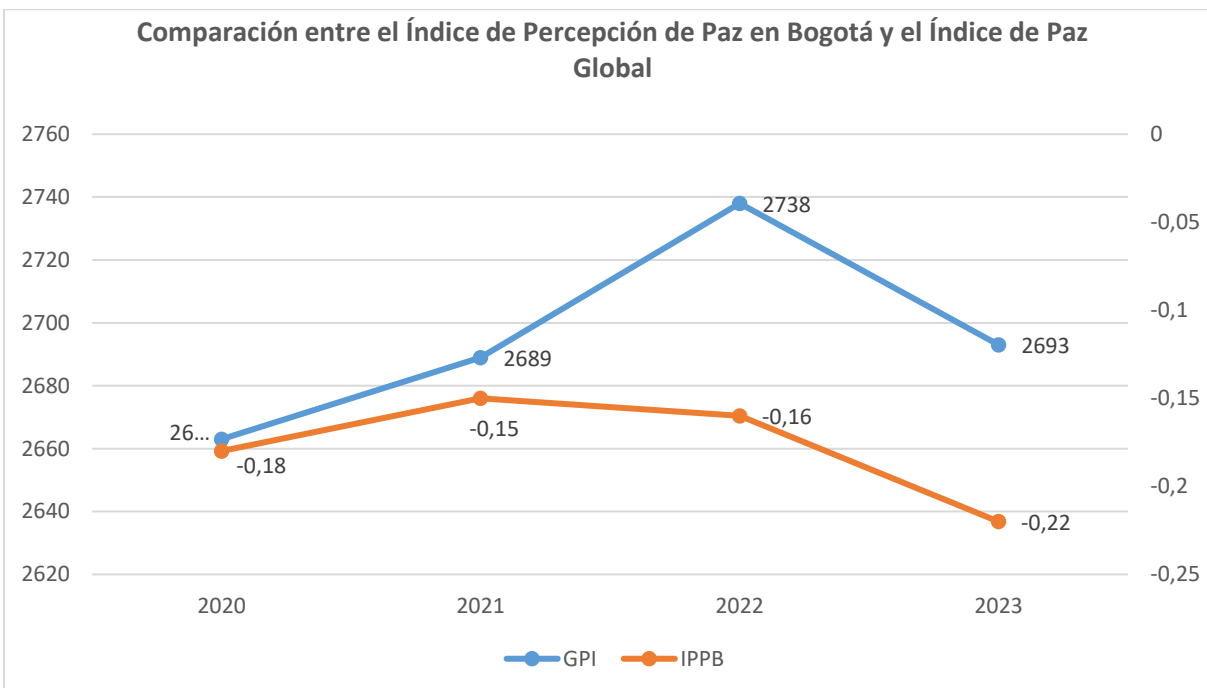
En conclusión, la línea de los acontecimientos del proceso de paz en Colombia entre 2020 y 2023, junto con los indicadores de seguridad ciudadana en Bogotá, ofrecen un marco valioso para comprender los resultados y el comportamiento del indicador de percepción de paz. Los avances y obstáculos en la implementación de los acuerdos de paz, como la reactivación de la violencia y la polarización política, se reflejan directamente en las fluctuaciones de los indicadores de seguridad en la ciudad. A medida que los homicidios y los hurtos aumentaron, la percepción de paz mostró una disminución significativa, lo que refuerza la correlación entre la seguridad ciudadana y la percepción social del proceso de paz. Así, los eventos nacionales y locales

contribuyen de manera integral al comportamiento del índice de percepción de paz, evidenciando la compleja interacción entre la política, la seguridad y las dinámicas sociales en Bogotá.

4.3.3 Comparación entre el Índice de Percepción de Paz en Bogotá y el Índice de Paz Global

La construcción del Índice de Percepción de Paz en Bogotá (IPPB) se fundamenta en un enfoque específico para medir la percepción de paz a nivel local, utilizando técnicas avanzadas de análisis de sentimientos y ponderaciones basadas en la popularidad de los medios de comunicación digitales. Este modelo se diferencia significativamente del Índice de Paz Global (IPG), cuyo alcance es global y se centra en una metodología estandarizada que combina indicadores cuantitativos y cualitativos relacionados con niveles de violencia, conflictos internos y externos, y grado de militarización de los países.

El IPPB se caracteriza por su naturaleza dinámica y ajustada al contexto local, proporcionando una representación matizada de la percepción ciudadana mediante un enfoque directo sobre el discurso mediático. Este índice integra publicaciones digitales clasificadas en categorías positivas, negativas y neutrales, y aplica una fórmula que considera tanto la proporción de publicaciones como su relevancia relativa según la popularidad del medio. De este modo, el IPPB refleja la prevalencia de narrativas en los medios. En contraste, el IPG utiliza una combinación de indicadores estáticos y globales que incluyen tasas de criminalidad, conflictos armados, acceso a armas y variables socioeconómicas para evaluar la paz a nivel nacional. Este enfoque holístico permite comparaciones internacionales y el análisis de tendencias macro, pero carece de la granularidad y adaptabilidad que ofrece el IPPB en su contexto local.



*Ilustración 28: Comparación entre el índice de percepción de paz en Bogotá y el índice de paz global
Elaboración por autores*

Como se evidencia en la imagen, el Índice de Paz Global (IPG) y el Índice de Percepción de Paz en Bogotá (IPPB) presentan tendencias diferenciadas a lo largo de los años, lo que refleja la influencia de los enfoques y metodologías utilizados en su construcción. Si bien ambos índices están diseñados para medir la paz, su naturaleza y alcance generan resultados que no son directamente comparables pero que permiten identificar patrones interesantes.

El IPG, calculado a nivel global, muestra una tendencia creciente entre 2020 y 2022, con valores que pasan de 2663 en 2020 a 2738 en 2022, lo que indica un deterioro progresivo de las condiciones de paz en Colombia según los indicadores globales. Sin embargo, en 2023, se observa una leve mejora con una reducción del índice a 2693, lo que sugiere un posible alivio en algunas dimensiones de la paz a nivel nacional, aunque el índice sigue siendo más alto que en 2020. Este comportamiento es típico de índices macro, que responden de manera más lenta a los cambios locales debido a la agregación de múltiples factores, como los conflictos armados, los gastos en seguridad y la militarización. Por otro lado, el IPPB, diseñado para medir percepciones locales mediante análisis de sentimientos en publicaciones digitales, presenta una tendencia más dinámica y directamente vinculada a la percepción de paz de los ciudadanos. En este índice, las percepciones comienzan en un nivel negativo de -0.18 en 2020, mejoran ligeramente en 2021 (-0.15) y 2022 (-0.16), pero muestran un deterioro marcado en 2023, alcanzando -0.22. Este retroceso podría estar relacionado con la percepción pública frente a eventos específicos, como la implementación de políticas de paz o los niveles de violencia en Bogotá, que afectan directamente la narrativa en los medios.

Ambos índices reflejan un deterioro en las condiciones de paz durante los últimos cuatro años, aunque lo hacen con distintos matices. Mientras que el IPG sugiere una mejora parcial en 2023, el IPPB indica un aumento en la percepción negativa durante el mismo periodo. Esta divergencia podría explicarse por el enfoque de cada índice: el IPPB, al basarse en discursos mediáticos, puede estar más influenciado por eventos recientes y coyunturales, como protestas, negociaciones fallidas o crisis de seguridad locales, mientras que el IPG responde a indicadores estructurales que tardan más en reflejar los cambios.

La comparación entre ambos índices pone en evidencia la complejidad de medir la paz y la importancia de considerar múltiples perspectivas. El IPG ofrece una visión más amplia y estable, útil para identificar tendencias generales en la paz de un país. En contraste, el IPPB proporciona una instantánea más ajustada a la percepción ciudadana, capturando las reacciones inmediatas a eventos locales. Esta complementariedad destaca la necesidad de interpretar ambos índices en conjunto, reconociendo sus limitaciones y fortalezas, para tener un entendimiento más completo de la dinámica de paz en Colombia. Integrar estas perspectivas permite relacionar tendencias globales con percepciones locales, identificando cómo los contextos estructurales y narrativos influyen en la percepción de paz de los ciudadanos.

5 CONCLUSIONES Y TRABAJOS FUTUROS

5.1 CONCLUSIONES

De acuerdo con la base de datos obtenida de los periódicos de Bogotá entre los años 2020 y 2022, los bogotanos asocian la paz con conceptos fundamentales como la libertad, la democracia y la justicia. En el contexto de una ciudad diversa y compleja como Bogotá, donde conviven distintas realidades sociales, económicas y culturales, estos términos adquieren una relevancia particular. La libertad se percibe como la capacidad de los ciudadanos para manifestarse sin restricciones en una ciudad que ha sido escenario de importantes movimientos sociales, la democracia implica el respeto por los derechos de todos los sectores y la participación en la toma de decisiones en un entorno urbano dinámico, y la justicia se considera clave para la resolución pacífica de los conflictos que surgen en este espacio de convivencia. Estas percepciones reflejan una visión donde la paz no se limita a la ausencia de violencia, sino que está profundamente ligada a la construcción de un entorno equitativo e inclusivo, esencial para el bienestar y la cohesión social en la capital colombiana.

El modelo base de BERT, en su variante BETO, requirió modificaciones sustanciales para ajustarse a la percepción de paz específica de los bogotanos. Aunque originalmente BERT clasifica el texto en categorías generales de sentimiento (positivo, neutral o negativo), en este caso fue necesario realizar un ajuste fino para capturar matices más profundos relacionados con conceptos clave como la democracia, la justicia y la libertad. Estos términos, fundamentales para la visión de paz en Bogotá, no podían ser abordados de manera efectiva sin adaptar el modelo a un enfoque que evaluara si los textos reflejan actitudes favorables o desfavorables hacia estas nociones. El proceso de ajuste fino permitió que el modelo se entrenara con un conjunto de datos específico, logrando así una mayor precisión en la clasificación de las emociones vinculadas con la percepción de paz de los ciudadanos de Bogotá durante los años 2020 al 2022.

Los resultados del reporte de clasificación mostraron un accuracy global de 0.74, lo que indica que el 74% de las predicciones realizadas por el modelo fueron correctas. En términos de precisión, el desempeño fue notablemente alto en la categoría Negativa, con un f1-score de 0.78, reflejando una alta eficacia en la identificación de sentimientos positivos en los artículos. Las categorías Positiva y Neutra alcanzaron f1-scores de 0.71 y 0.70, respectivamente, lo que denota un desempeño balanceado del modelo en estas categorías. Estos resultados demuestran la robustez del modelo ajustado y su capacidad para interpretar los matices emocionales presentes en los textos analizados.

El análisis de los artículos de 2023 reveló una tendencia significativa hacia una cobertura negativa del tema de la paz en Bogotá. Medios como *Semana* y *El Espectador* destacaron por una alta proporción de artículos con sentimiento negativo (78.8% y 67.1%, respectivamente), mientras que *Portafolio* también presentó un predominio de artículos negativos (52.2%), aunque menos acentuado. En contraste, *La República* y *2 Orillas* ofrecieron una cobertura más equilibrada, con una mayor proporción de artículos neutros y positivos. En conjunto, los cinco periódicos evaluados

evidenciaron que el 56.1% de las publicaciones sobre la paz en 2023 tenían un enfoque negativo, lo que podría influir desfavorablemente en la percepción pública sobre los avances hacia la paz en Bogotá.

Finalmente, el análisis del Índice de Percepción de Paz en Bogotá revela algunos resultados clave que permiten comprender mejor la evolución de la percepción pública sobre la paz en la ciudad. El indicador, al integrar las publicaciones positivas, negativas y neutrales, demuestra cómo las fluctuaciones en la política de paz y la dinámica social inciden en el sentimiento colectivo de la población. Durante los años analizados, se observaron variaciones que reflejan las tensiones políticas y sociales del país, especialmente en torno a las políticas de paz implementadas por el gobierno. Los resultados muestran una tendencia a la baja en la percepción de paz durante 2022 y 2023, lo cual puede explicarse por el aumento de la violencia y la polarización política, que influyó en un mayor volumen de publicaciones negativas, mientras que las publicaciones positivas fueron incapaces de contrarrestar este impacto. La inclusión de publicaciones neutrales, que representan incertidumbres o falta de una postura clara, permitió que el indicador no se viera excesivamente influenciado por los picos emocionales, brindando una medición más equilibrada y objetiva. Estos resultados sugieren que, aunque se han dado pasos hacia la paz, la percepción pública sigue siendo vulnerable a los eventos violentos y a la polarización política, lo que resalta la necesidad de un enfoque integral y sostenido en la implementación de políticas de paz para generar una percepción más estable y positiva en la población de Bogotá.

5.2 TRABAJOS FUTUROS

En el marco de esta investigación, se propone realizar un análisis adicional utilizando un enfoque más amplio que contemple una base de datos compuesta por otras clases de publicaciones digitales, específicamente tweets. Este tipo de datos, caracterizados por su brevedad y concisión, ofrece una perspectiva diferente en comparación con las publicaciones de los periódicos digitales. La inclusión de tweets permitirá explorar la percepción pública de la paz en un formato más dinámico, lo que podría enriquecer los hallazgos de este estudio. Al analizar un conjunto diverso de datos provenientes de redes sociales, se podrá capturar la evolución de las opiniones y sentimientos sobre temas relevantes de manera más inmediata. Este enfoque no solo facilitará una comparación entre las distintas plataformas de publicación, sino que también contribuirá a la construcción de un modelo analítico más robusto y representativo de las discusiones contemporáneas en torno a la paz.

Una propuesta clave para futuros trabajos es ampliar el círculo de implementación del modelo, extendiendo el análisis más allá de Bogotá y abarcando datos de todo el territorio colombiano, así como de otros países. Esta ampliación permitirá una comprensión más integral de la percepción de la paz en diferentes contextos regionales e internacionales, lo que es esencial dado el contexto diverso y complejo que caracteriza a cada una de las naciones. El objetivo de esta ampliación es generar una herramienta analítica que no solo se limite a evaluar las dinámicas locales, sino que también se compare con índices de paz utilizados actualmente. Al integrar datos de múltiples regiones y naciones, se podrán identificar patrones y tendencias que enriquezcan el análisis, ofreciendo una visión más completa de la situación global en términos de paz.

Además, esta herramienta podría servir como un recurso valioso para corregir o incorporar elementos en los índices de paz existentes, aportando datos empíricos que ayuden a refinar las métricas actuales y faciliten la toma de decisiones en políticas públicas relacionadas con la paz en diversas realidades geográficas.

6 REFERENCIAS BIBLIOGRÁFICAS

- [1] F. Harto de Vera. “La construcción del concepto de paz: Paz negativa, paz positiva y paz imperfecta”. Dialnet. Accedido el 14 de noviembre de 2023. [En línea]. Disponible: <https://dialnet.unirioja.es/servlet/articulo?codigo=5832796>
- [2] “Índice de paz global”. Datosmacro.com. Accedido el 14 de noviembre de 2023. [En línea]. Disponible: <https://datosmacro.expansion.com/demografia/indice-paz-global>
- [3] Peace Research Institute Oslo (PRIO), Universidad de los Andes (2018) “Percepciones y Experiencias relacionadas con el Proceso de Paz en Colombia”.
- [4] Gualy, L. F. (2014). Construcción de cultura de paz en América Latina desde la educación superior. Revista Universidad de La Salle, 65.
- [5] Hernández, D. A. (2019). Nociones de paz: una revisión teórica del concepto. Ciudad Paz-Ando, 12(1). <https://doi.org/10.14483/2422278x.13951>
- [6] A. Calvo. “Breve historia de la paz (II): Edad Media y Moderna” (2017). speculorum. [En línea]. Disponible: <https://speculorum.wordpress.com/2017/11/10/breve-historia-de-la-paz-ii-edad-media-y-moderna/>
- [7] Tah Ayala, E. D. (2018). Las Relaciones Internacionales desde la perspectiva social. La visión del Constructivismo para explicar la identidad nacional. Revista Mexicana de Ciencias Políticas y Sociales, 63(233). <https://doi.org/10.22201/fcpys.2448492xe.2018.233.62593>
- [8] Grupo de expertos designado por el secretario general de las Naciones Unidas. (2004, octubre 1221). Un Mundo más Seguro: Nuestra Responsabilidad Compartida. Naciones Unidas. <https://documents.un.org/doc/undoc/gen/n04/602/34/pdf/n0460234.pdf?token=CYhkyKlf3wBrwjSMtf&fe=true>
- [9] M, Romero (2023) Percepciones de paz y conflicto armado en actores escolares
- [10] S, Céspedes (2019) “Proyecto Construcción de Zonas de Paz: análisis de la percepción de la niñez y la adolescencia sobre la violencia y la paz en la comunidad de Guararí” <https://www.mendeley.com/reference-manager/reader-v2/47717adb-b5f3-3875-a660-3a5547287734/18bf8fc1-121a-e0f7-7cd3-8d78b58cd61d/>
- [11] Del Pozo Serrano, F., Jiménez Bautista, F. & Barrientos Soto, A., (2022) Pedagogía Social y Educación Social en Colombia: cómo construir la cultura de paz comunitaria en el postconflicto
- [12] V, Torres (2021) La paz territorial como enfoque para comprender la justicia transicional y la construcción de paz en Colombia: mirada desde el Distrito Especial de Santiago de Cali
- [13] M. Sánchez (2023) Educación y construcción de paz <https://www.mendeley.com/reference->

manager/reader-v2/ff33ce7c-028c-3085-84ec-26e73bb1f631/66f196ae-24e9-8ab9-32d8-dcd9f5aec449

[14] Datos macro.com. (2022). Índice de Paz Global 2022. Índice de Paz Global. <https://datosmacro.expansion.com/demografia/indice-paz-global>

[15] K. Birke Daniels y S. Kurtenbach, Eds. (2021), Los enredos de la paz Reflexiones alrededor del largo camino de la transformación del conflicto armado en Colombia. Bogotá: Friedrich-Ebert-Stiftung en Colombia (FESCOL), German Inst. Global Área Stud. (GIGA), Disponible: <https://library.fes.de/pdf-files/bueros/kolumbien/18212-20210901.pdf>

[16] V. M. OROZCO PUELLO (2019) "Desarrollo de un prototipo de aplicación web que permita la extracción de las ofertas laborales de las principales plataformas que postulan empleos en la región caribe, usando la técnica web SCRAPING.", Trabajo de grado para el título de INGENIERO DE SISTEMAS, UNIV. DEL SINU ELIAS BECHARA ZAINUM SECC. CARTAGENA
Disponible: <http://repositorio.unisinucartagena.edu.co:8080/xmlui/handle/123456789/94>

[17] Meléndez Gómez N. Briceño M. (2023) MINERÍA DE TEXTO APLICADO AL ANÁLISIS DE SENTIMIENTOS PARA LA BÚSQUEDA DE PATRONES DE MOTIVACIÓN EN PROGRAMA DE ACTUALIZACIÓN DOCENTE

[18] Atkinson-Abutridy, J. (2022). Text analytics: An introduction to the science and applications of unstructured information analysis. In Text Analytics: An Introduction to the Science and Applications of Unstructured Information Analysis. <https://doi.org/10.1201/9781003280996>

[19] Amat, J (2020) Análisis de texto (text mining) con Python.

[20] Dipanjan Sarkar (2019) Text Analytics with Python: A Practitioner's Guide to Natural Language Processing.

[21] Giraldo A, Orozco Duque F (2023) Evolución del procesamiento natural del lenguaje

[22] Rodríguez D (2024) "Calcular la similitud de Jaccard en Python", Analytics Lane. Disponible: <https://www.analyticslane.com/2024/01/19/calcular-la-similitud-de-jaccard-en-python/>.

[23] R. Campos, V. Mangaravite, A. Pasquali, A. Jatowt, A. Jorge, C. Nunes, and A. Jatowt, "YAKE! Keyword extraction from single documents using multiple local features," Information Sciences, vol. 509, pp. 257–289, 2020.

[24] Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. In Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. <https://doi.org/10.1017/CBO9781139084789>

[25] B, Abayomi (2023) A BERT Framework to Sentiment Analysis of Tweets

[26] Zhang, Y., Tang, R., & Lin, J.J. (2019). Explicit Pairwise Word Interaction Modeling Improves Pretrained Transformers for English Semantic Similarity Tasks. ArXiv, abs/1911.02847.

[27] Departamento nacional de planeación -DNP, Guía metodológica para la formulación de indicadores. Bogotá: Scripto Gomez Rosales Asoc. Cia. Ltda., 2009. Disponible: <https://colaboracion.dnp.gov.co/CDT/Inversiones%20y%20finanzas%20pblicas/Guia%20Metodologica%20Formulacion%20-%202010.pdf>

[28] Departamento nacional de planeación -DNP, Guía para la construcción y análisis de indicadores. Bogotá, 2018. Accedido el 9 de noviembre de 2024. Disponible: https://colaboracion.dnp.gov.co/CDT/Sinergia/Documentos/Guia_para_elaborar_Indicadores.pdf

[29] Consejo Nacional de Evaluación de la Política de Desarrollo Social - CONEVAL, Manual para el diseño y la construcción de indicadores. Instrumentos principales para el monitoreo de programas sociales de México. México D.F., 2013. Accedido el 6 de noviembre de 2024. Disponible: https://www.coneval.org.mx/Informes/Coordinacion/Publicaciones%20oficiales/MANUAL_PARA_EL_DISENO_Y_CONTRUCCION_DE_INDICADORES.pdf

[30] J. C. Sobrino Sande, "Análisis de sentimiento en Twitter", Máster universitario en ingeniería informática, Univ. Oberta Catalunya, Barcelona, España, 2018. Accedido el 22 de noviembre de 2023. Disponible: <http://hdl.handle.net/10609/81435>

[31] Entidad pública, INSTITUTO COLOMBIANO DE CRÉDITO EDUCATIVO Y ESTUDIOS TÉCNICOS EN EL EXTERIOR "MARIANO OSPINA PÉREZ" "Modelo de análisis de sentimientos en redes sociales con un modelo de minería de datos y machine learning". Accedido el 24 de noviembre de 2023. [En línea]. Disponible: <https://herramientas.datos.gov.co/usos/modelo-de-analisis-de-sentimientos-en-redes-sociales-con-un-modelo-de-mineria-de-datos-y>

[32] C. Hernández, G. Chanchí "Análisis de sentimientos sobre la percepción de los participantes de una feria virtual de emprendimiento durante el confinamiento". Accedido el 24 de noviembre de 2023.

[33] A. Özberk and İ. Çiçekli, "Offensive Language Detection in Turkish Tweets with Bert Models," 2021 6th International Conference on Computer Science and Engineering (UBMK), Ankara, Turkey, 2021, pp. 517-521, doi: 10.1109/UBMK52708.2021.9559000.

[34] L. S. Liebovitch et al., "Complexity Analysis of Sustainable Peace: Mathematical Models and Data Science Measurements," New Journal of Physics, vol. 21, art. 073022, pp. 1-10, Jul. 2019. Disponible: <https://doi.org/10.1088/1367-2630/ab2a04>

[35] A. Pozo, "Measuring Peace. Initiatives, Limitations and Proposals", en Barcelona, España, 4-5 de marzo de 2010. Barcelona: Inst. Catala Int. Per Pau, 2010, p. 63. Accedido el 11 de noviembre de 2024.

Disponible: https://www.icip.cat/wp-content/uploads/2020/11/icip_documento_1_03a_2010.pdf

[36] SCImago Lab (2022) "Ranking de SCImago para la Reputación Web de Medios Digitales". https://www.scimagomedia.com/rankings.php?region=Latin%20America&country=Colombia&edition=2022_12

[37] Newtral (2021). "Acuerdo de Paz en Colombia: cinco años en cinco datos". Disponible: El acuerdo de paz en Colombia: cinco años en cinco datos

[38] Infobae (2022). "Durante 2021 en Colombia se recrudeció el conflicto armado y la violencia, asegura la Cruz Roja". Disponible: Durante 2021 en Colombia se recrudeció el conflicto armado y la violencia, asegura la Cruz Roja - Infobae

[39] Fundación Ideas para la Paz, El camino de la Paz Total 2024. Bogotá, Colombia: Fundación Ideas para la Paz, 2024. [En línea]. Disponible: <https://globalinitiative.net/wp-content/uploads/2023/08/El-Camino-de-la-Paz-Total.pdf>.

[40] Fundación para el Progreso de la Región PROBOGOTÁ, Informe Anual de Seguridad 2023. Bogotá, Colombia: PROBOGOTÁ, 2024.

7 ANEXO 1

A continuación, se comparte el enlace al repositorio del proyecto, donde se pueden consultar todos los archivos, códigos y recursos utilizados para la realización del análisis. Este repositorio contiene el material relevante para reproducir los resultados, así como la documentación detallada del proceso metodológico y los hallazgos obtenidos. El acceso al repositorio permite explorar más a fondo los datos, así como realizar modificaciones o extensiones al proyecto.

<https://github.com/MelanieGazabon/Proyecto-Paz.git>