

**Evaluación de la percepción ciudadana en la red social X mediante técnicas de minería y
analítica de datos para el fortalecimiento institucional de la Secretaría Distrital de
Hacienda**

John Jairo Moreno Bohórquez

Nicolás Rodríguez Medina

Asesor

Rafael Gaitán Ospina

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica

2025

Resumen

Este trabajo analiza la percepción ciudadana hacia la Secretaría Distrital de Hacienda mediante el estudio de publicaciones en la red social X, utilizando técnicas de análisis de sentimientos para comprender cómo reaccionan los usuarios frente a sus políticas y acciones institucionales. La problemática se centra en las limitaciones que presentan los modelos automáticos para interpretar lenguaje informal, sarcasmo y las particularidades del español usado en redes sociales. Para abordar este proyecto, se realizó la recolección, limpieza y preparación de los datos, junto con la aplicación de un modelo de clasificación de sentimientos y un análisis exploratorio que permitió identificar patrones discursivos, temas recurrentes y tendencias de conversación. Los resultados evidencian picos de actividad asociados a eventos específicos, una predominancia de mensajes neutrales y una menor presencia de contenido positivo y negativo, lo que refleja tanto el comportamiento natural de la conversación digital como las limitaciones del modelo empleado. Estos hallazgos ofrecen insumos relevantes para comprender las principales preocupaciones ciudadanas y aportan elementos que pueden fortalecer la comunicación institucional y la toma de decisiones basada en datos.

Palabras clave: Análisis, Sentimientos, Percepción, Ciudadanía, Decisiones

Abstract

This study analyzes public perception of the District Treasury Secretariat by examining posts on social network X, using sentiment analysis techniques to understand how users react to its policies and institutional actions. The problem lies in the limitations of automated models in interpreting informal language, sarcasm, and the specific characteristics of Spanish used on social media. To address this issue, data was collected, cleaned, and prepared, and a sentiment classification model was applied, along with an exploratory analysis that identified discursive patterns, recurring themes, and conversation trends. The results show activity spikes associated with specific events, a predominance of neutral messages, and a lower presence of positive and negative content, reflecting both the natural behavior of digital conversation and the limitations of the model used. These findings offer relevant insights for understanding the main concerns of citizens and provide elements that can strengthen institutional communication and data-driven decision-making.

Keywords: Analysis, Feelings, Perception, Citizenship, Decisions

Tabla de Contenido

Introducción	6
Justificación	8
Objetivos.....	10
Objetivo General.....	10
Objetivos Específicos	10
Marco de Referencia	11
Marco Conceptual.....	11
Antecedentes Relevantes	13
Metodología	14
Diagrama de Fases de Implementación	17
Fase 1 Limpieza y Normalización de los Datos.....	18
Fase 2 Clasificación de Sentimientos	23
Metodología y Coherencia con el Enfoque del Proyecto	23
Transparencia y Facilidad de Interpretación	23
Eficiencia Computacional y Viabilidad Operativa	24
Fase 3: Análisis Temático y Visualización de Resultados.....	30
Conclusiones.....	41
Recomendaciones	43
Referencias.....	44

Lista de Figuras

Figura 1 <i>Diagrama de Fases de Implementación</i>	17
Figura 2 <i>Tabla Exploratoria de los Primeros Registros del Dataset de Tweets</i>	18
Figura 3 <i>Función de Limpieza de Texto para Tweets en Python</i>	19
Figura 4 <i>Proceso de Limpieza y Normalización del Texto en el Análisis de Tweets</i>	20
Figura 5 <i>Análisis de Sentimientos de Tweets en la Red Social X con TextBlod</i>	24
Figura 6 <i>Evaluación de Métricas de Clasificación de Sentimientos</i>	27
Figura 7 <i>Mapa de Calor con Métricas de Desempeño por Clase en Clasificación de Sentimientos</i>	29
Figura 8 <i>Top 10 de Usuarios Más Activos en la Conversación Digital Sobre la SDH</i>	32
Figura 9 <i>Nube de Palabras sobre la Secretaría Distrital de Hacienda (X)</i>	34
Figura 10 <i>Dispersión Temporal de Sentimientos en Tweets</i>	35
Figura 11 <i>Evolución Temporal de Sentimientos Negativos en la Red Social X</i>	36
Figura 12 <i>Usuarios con Mayor Actividad en X Según Conteo de Publicaciones</i>	37
Figura 13 <i>Conteo Diario de Tweets Sobre la Secretaría Distrital de Hacienda</i>	38
Figura 14 <i>Distribución Porcentual de Mensajes por Polaridad en X</i>	39
Figura 15 <i>Detalle Tabular de Tweets Analizados en Estudio de Sentimientos</i>	40

Introducción

En un entorno marcado por transformaciones constantes en los ámbitos social, tecnológico y económico, los ciudadanos recurren a la red social X (antes Twitter) como medio para expresar sus opiniones, emociones y percepciones frente a temas de interés público. Esta plataforma se ha consolidado como una herramienta clave para comprender las dinámicas de la percepción ciudadana. En este contexto, el presente trabajo se enfoca en evaluar los sentimientos y opiniones manifestados por los usuarios en torno a la Secretaría Distrital de Hacienda, con el propósito de identificar las principales preocupaciones, temáticas recurrentes y emociones asociadas a su gestión institucional.

El enfoque que se adoptó combina fundamentos teóricos sobre el análisis de sentimientos y la comunicación digital, con la aplicación de técnicas de analítica de datos sobre publicaciones reales. Se incluye también una mirada crítica sobre la utilidad de estas herramientas para la gestión pública, especialmente en lo relacionado con la transparencia, la toma de decisiones y la mejora del servicio al ciudadano.

En el desarrollo del trabajo se describirá el enfoque metodológico que se aplicará para analizar los mensajes recolectados, detallando las etapas de preparación de datos, procesamiento analítico y visualización de resultados. Asimismo, se explicará cómo cada fase aporta a la comprensión de la percepción ciudadana en la red social X y cómo los hallazgos esperados pueden apoyar la toma de decisiones institucionales. En lugar de centrarse en conceptos teóricos, esta sección expondrá los procedimientos, herramientas y criterios que guiarán el análisis, así como el tipo de información que se podrá obtener a partir del procesamiento de los mensajes ciudadanos durante el periodo seleccionado.

Finalmente, el documento presentará conclusiones directamente relacionadas con los objetivos planteados y con el proceso desarrollado a lo largo de cada capítulo. En estas conclusiones se sintetizará cómo, a partir de la recolección de mensajes en la red social X, se llevó a cabo su limpieza y preparación, se aplicaron técnicas de análisis exploratorio y se construyeron visualizaciones que permitieron identificar patrones discursivos, tendencias emocionales y temas recurrentes asociados a la Secretaría Distrital de Hacienda.

Asimismo, se explicará de manera puntual cómo cada etapa metodológica desde la depuración de los datos, el análisis descriptivo del contenido, la evaluación del comportamiento lingüístico de los usuarios y la elaboración de tableros de control contribuyó al logro de los objetivos del proyecto. Estas conclusiones reflejarán los principales hallazgos derivados del análisis, destacando los comportamientos ciudadanos observados, las temáticas predominantes y las variaciones en la conversación digital durante el periodo estudiado.

De igual forma, se expondrán recomendaciones basadas en los resultados obtenidos y se señalarán posibles líneas de investigación futura, especialmente en lo relacionado con el uso de datos provenientes de redes sociales para fortalecer la comunicación institucional, mejorar los procesos de análisis de opinión pública y apoyar la toma de decisiones fundamentadas en evidencia.

Justificación

La elección de esta solución tecnológica responde a la necesidad de comprender con mayor profundidad la percepción ciudadana frente al desempeño institucional en entornos digitales, un fenómeno que en la actualidad tiene un impacto significativo en la manera en que las entidades públicas comunican, rinden cuentas y ajustan sus estrategias. El análisis de sentimientos aplicado a redes sociales, y en particular a los mensajes emitidos en X hacia la cuenta oficial de la Secretaría de Hacienda Distrital, representa no solo un desafío técnico y metodológico, sino también una herramienta clave con implicaciones prácticas para la toma de decisiones, la mejora continua y el fortalecimiento de la relación entre el Estado y la ciudadanía.

En la Secretaría Distrital de Hacienda no existen soluciones tecnológicas que permitan procesar de manera masiva y en tiempo real las opiniones de los ciudadanos sobre campañas institucionales y servicios ofrecidos. Esta limitación genera vacíos en los mecanismos de retroalimentación y dificulta una lectura oportuna y sistemática de las expectativas sociales. Frente a esta necesidad, la solución planteada busca aportar información fundamentada en datos reales mediante la extracción automática de tweets, el tratamiento y limpieza de los datos, y la aplicación de algoritmos de análisis de sentimientos que asignan valoraciones cuantitativas y cualitativas a cada mensaje procesado. Además, se incorporan diversas herramientas de visualización que incluyen distribuciones de sentimientos, nubes de palabras representativas y listados detallados de mensajes, las cuales facilitan la interpretación de los patrones identificados en la conversación digital. De igual manera, los datos procesados se almacenan de forma estructurada en una base de datos MySQL, permitiendo mantener un registro histórico y realizar comparaciones temporales que fortalecen el análisis y el seguimiento continuo de la percepción ciudadana.

Por lo tanto, no realizar este proyecto implicaría mantener las limitaciones actuales en cuanto a monitoreo de opinión pública, prolongando la dependencia de métodos tradicionales como encuestas manuales o análisis no sistematizados. Esto afectaría la capacidad de la entidad para adaptar sus estrategias institucionales de manera ágil, limitaría la identificación temprana de problemas o demandas ciudadanas, y restringiría la generación de evidencia de datos reales útiles para la toma de decisiones. Asimismo, se desaprovecharían las oportunidades que ofrece la inteligencia artificial y el procesamiento de lenguaje natural como herramientas clave para la innovación pública y el fortalecimiento de la relación entre el estado y la ciudadanía. Adicionalmente, al integrarse en una plataforma web accesible, esta herramienta garantiza su usabilidad práctica y su proyección sostenible en diferentes contextos administrativos.

Objetivos

Objetivo General

Descubrir patrones de opiniones positivas, negativas y neutras expresadas por los ciudadanos en la red social X sobre la gestión de la Secretaría Distrital de Hacienda, empleando técnicas de analítica de datos.

Objetivos Específicos

Aplicar técnicas de limpieza y normalización de texto incluyendo eliminación de caracteres no relevantes, emojis, signos de puntuación, URLs y palabras vacías para preparar los datos y asegurar su calidad semántica.

Clasificar los mensajes recolectados en categorías de sentimiento (positivo, negativo y neutro) utilizando técnicas de análisis de sentimientos y herramientas de análisis de texto, con el fin de obtener datos estructurados que faciliten la interpretación de la percepción ciudadana.

Evaluar los resultados de la minería de texto para identificar los temas más relevantes y priorizar las áreas de gestión de la Secretaría Distrital de Hacienda que generan mayor interés o inconformidad ciudadana.

Marco de Referencia

Marco Conceptual

El análisis de sentimientos, también denominado minería de opiniones, es una técnica del Procesamiento de Lenguaje Natural (PLN), campo de la inteligencia artificial que se encarga de estudiar y desarrollar métodos para que las máquinas puedan comprender, interpretar y generar lenguaje humano de manera significativa. Esta disciplina integra modelos estadísticos, lingüísticos y de aprendizaje automático para extraer información útil de textos. En este contexto, el análisis de sentimientos se orienta a identificar emociones, percepciones y valoraciones subjetivas en contenidos digitales. Su importancia ha aumentado especialmente en el estudio de conversaciones públicas en redes sociales, donde los usuarios expresan opiniones de forma espontánea, masiva y en tiempo real (Carangui et al., 2024).

La red social X (antes Twitter) se ha consolidado como una de las fuentes más utilizadas para estos fines, debido a su carácter inmediato, público y de alto volumen informativo. Desde los estudios fundacionales de Pak y Paroubek (2010), quienes establecieron a Twitter como un corpus válido para tareas de análisis de sentimiento, múltiples investigaciones han demostrado su utilidad para comprender dinámicas sociales complejas. Por ejemplo, Tumasjan et al. (2010) evidenciaron que los mensajes publicados en esta plataforma pueden reflejar tendencias políticas y anticipar resultados electorales.

Otros trabajos, como el de Nakov et al. (2013) en SemEval, han contribuido a estandarizar tareas, métricas y corpus para la evaluación comparativa de modelos de sentimiento en Twitter. Asimismo, estudios contemporáneos han mostrado avances en el uso de técnicas más robustas, como redes neuronales profundas para detectar polaridad y sesgos en contenidos

políticos y mediáticos (Yenkikar et al., 2022) o para identificar estados emocionales asociados a condiciones como la depresión (Babu & Kanaga, 2022).

En el ámbito estratégico institucional, Carangui et al. (2024) demuestran que la aplicación del análisis de sentimientos mediante inteligencia artificial fortalece la gestión comunicacional, permitiendo monitorear la reputación digital y adaptar los mensajes según la reacción de las audiencias. En esta misma línea, Mostafa (2013) muestra cómo la minería de texto en redes sociales ayuda a comprender percepciones ciudadanas hacia marcas y servicios, reforzando la utilidad del análisis emocional como insumo para la toma de decisiones basadas en evidencia.

Este proyecto se fundamenta en estos aportes para diseñar una solución tecnológica basada en software libre que permita extraer, procesar y analizar publicaciones relacionadas con la Secretaría Distrital de Hacienda, generando visualizaciones que apoyen la interpretación institucional de la percepción ciudadana. Se integra además el enfoque metodológico planteado por Valencia et al. (2024), quienes proponen modelos basados en minería de texto y diccionarios de sentimientos para identificar patrones emocionales en temáticas públicas, como el transporte en América Latina.

Finalmente, investigaciones recientes en plataformas como Instagram (García, 2024), YouTube y Twitter (García Merayo et al., 2023), así como revisiones generales accesibles en repositorios académicos (EBSCO, s. f.), destacan que el análisis de sentimientos es un campo interdisciplinar en expansión, que integra PLN, minería de texto y aprendizaje automático para interpretar fenómenos sociales, políticos y comunicacionales.

Antecedentes Relevantes

Diversos estudios internacionales han evidenciado que el análisis de sentimientos permite comprender no solo emociones individuales, sino también tendencias colectivas y su evolución en el tiempo. Por ejemplo, la investigación de Valencia et al. (2024) sobre transporte público en América Latina, basada en más de 28 000 comentarios, mostró una transición gradual de emociones negativas a positivas, reflejando mejoras percibidas en el servicio. Este tipo de análisis demuestra el valor del monitoreo emocional para la toma de decisiones en políticas públicas.

En contextos más mediáticos, García Merayo et al. (2023) demostraron que el uso de técnicas de minería de texto en redes sociales permite identificar cambios en la opinión pública frente a casos de alto impacto. De manera similar, Busón Buesa (2020) analizó cómo la minería de opinión permite caracterizar discursos de odio en redes sociales, subrayando la importancia de comprender las dinámicas emocionales en entornos digitales.

En el ecosistema de la red social X, múltiples estudios han comparado la efectividad de modelos para clasificación emocional, destacando diferencias metodológicas y de desempeño (Lovera & Cardinale, 2023). Además, trabajos como el de Reyes & Moreno (2024) sistematizan la aplicación de técnicas de PLN y machine learning para el análisis de sentimientos en ambientes reales, describiendo su pertinencia para modernizar sistemas de información organizacional.

La literatura consultada (TFG María Belén Alarcón Ramos, s. f.; EBSCO, s. f.) confirma que el análisis de sentimientos se ha convertido en una herramienta consolidada tanto en investigación académica como en aplicaciones institucionales, especialmente en situaciones donde la percepción ciudadana es un indicador crítico de desempeño, confianza y legitimidad.

Metodología

La metodología propuesta para el desarrollo del proyecto se estructuró bajo los lineamientos del Project Management Body of Knowledge (PMBOK), reconocido como uno de los estándares más sólidos y adoptados internacionalmente para la gestión eficiente de proyectos. Este marco metodológico permitió organizar el trabajo de manera sistemática a través de los procesos de iniciación, planificación, ejecución, monitoreo y cierre, garantizando la adecuada coordinación de las actividades, la calidad de los entregables y el cumplimiento de los objetivos estratégicos. Gracias a su enfoque integral, PMBOK facilitó una administración clara del alcance, tiempo, calidad, riesgos y recursos, asegurando que cada fase del proyecto se ejecute con claridad, control y trazabilidad.

Durante la etapa de planificación, se definirá detalladamente el alcance del proyecto, delimitando los objetivos relacionados con la recolección, limpieza, análisis y visualización de datos provenientes de la red social X. Asimismo, se identificarán los actores clave, los requerimientos funcionales y las restricciones técnicas que orientarán el desarrollo posterior. Esta fase permitirá establecer un cronograma estructurado, asignar responsabilidades y anticipar los riesgos asociados tanto al tratamiento de datos no estructurados como al análisis automatizado de sentimientos. El uso de herramientas de seguimiento y control permitirá mantener visibilidad permanente sobre el avance y aplicar acciones correctivas cuando sea necesario.

Bajo esta estructura metodológica, el proyecto se desarrollará en tres fases principales, cada una alineada con los objetivos específicos planteados. La primera fase se centrará en la limpieza, normalización y preparación del texto, etapa crucial para garantizar la calidad del análisis. En esta fase se realizará la depuración de caracteres especiales, la eliminación de ruido digital, la estandarización del formato de los mensajes y la aplicación de técnicas de

preprocesamiento como tokenización y eliminación de stopwords. Este proceso transformará los mensajes ciudadanos en un conjunto de datos limpio, coherente y adecuado para su procesamiento analítico. Su correcta ejecución será determinante, ya que la calidad de la información influirá directamente en la precisión del modelo de clasificación posterior.

La segunda fase consistirá en la clasificación de los mensajes por sentimiento, apoyándose en librerías especializadas de procesamiento del lenguaje natural. En esta etapa se evaluarán diferentes enfoques de análisis de sentimiento, se entrenará y ajustará el modelo seleccionado y se validará su precisión mediante métricas como precisión, recall y F1-score. La aplicación del modelo permitirá identificar patrones emocionales y tendencias en los comentarios ciudadanos, proporcionando una visión clara sobre la percepción pública frente a los temas tributarios y financieros abordados por la Secretaría Distrital de Hacienda. Asimismo, se documentarán las limitaciones del modelo y se implementarán controles para mitigar inconsistencias derivadas del lenguaje informal propio de las redes sociales.

La tercera fase estará enfocada en la visualización analítica de los resultados mediante tableros de control interactivos. Se desarrollarán dashboards que integrarán gráficas comparativas, histogramas de sentimientos, nubes de palabras que evidenciarán el desempeño del modelo. Esta fase permitirá consolidar los resultados en una herramienta útil para la toma de decisiones, facilitando la interpretación por parte de los equipos directivos y operativos. El diseño de los tableros priorizará la claridad visual, la navegabilidad y la generación ágil de informes estratégicos.

El uso del marco PMBOK ofrecerá un soporte integral al proceso metodológico no solo desde el punto de vista técnico, sino también desde la perspectiva organizacional. Dado que el análisis involucrará información sensible proveniente de ciudadanos, será indispensable

garantizar el cumplimiento de lineamientos institucionales, protocolos de manejo de datos y criterios de transparencia. La adecuada gestión documental, la trazabilidad de decisiones y el registro continuo de actividades asegurarán la coherencia entre los resultados esperados y los objetivos estratégicos de la entidad.

Asimismo, la metodología permitirá fortalecer la articulación secuencial de las fases y esto facilitará avanzar de manera ordenada desde la captura de datos no estructurados hasta la generación de conocimiento accionable. Cada etapa se alimentará de la anterior, construyendo una ruta metodológica sólida que permitirá un análisis confiable y alineado con la necesidad institucional de comprender la percepción ciudadana.

En este sentido, la aplicación del PMBOK no solo garantizará una ejecución disciplinada y estructurada, sino que permitirá anticipar riesgos técnicos como variaciones en el volumen de datos, sesgos semánticos o inconsistencias en la clasificación automatizada. La gestión proactiva de estos riesgos evitará retrasos y asegurará la calidad del producto final. A su vez, la metodología fomentará una mejora continua durante el desarrollo del proyecto, impulsando ajustes basados en hallazgos parciales y promoviendo un enfoque adaptativo orientado a la optimización del análisis.

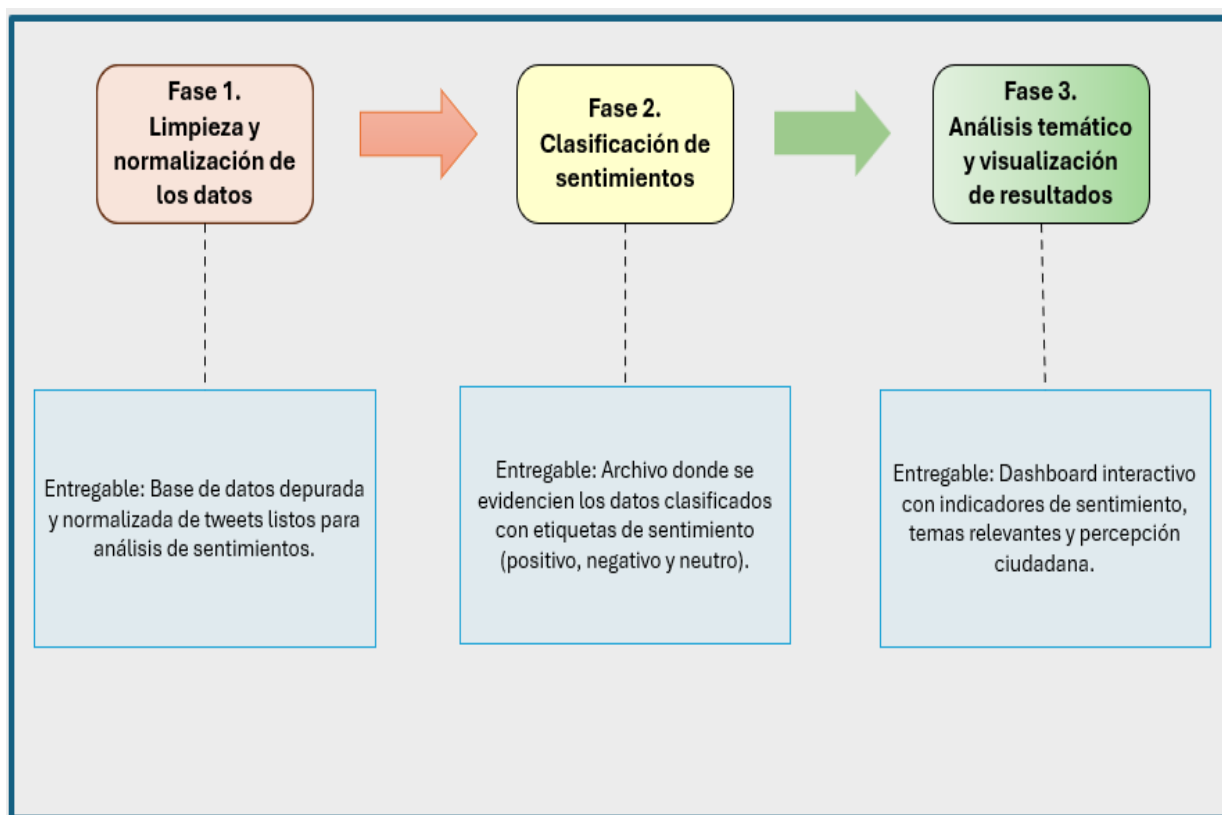
Finalmente, el proyecto permitirá consolidar un proceso analítico capaz de transformar los mensajes ciudadanos en información estructurada, interpretable y útil para la gestión institucional. El enfoque metodológico adoptado integrará técnicas de analítica de datos con prácticas de gestión de proyectos, generando un modelo replicable para futuros análisis y contribuyendo al fortalecimiento de la participación ciudadana, la transparencia institucional y la toma de decisiones basadas en evidencia dentro de la Secretaría Distrital de Hacienda.

Diagrama de Fases de Implementación

El diagrama de fases de implementación representa de manera sintética y estructurada la secuencia metodológica seguida durante el desarrollo del proyecto, permitiendo visualizar la relación lógica y progresiva entre cada una de las etapas del análisis. Este esquema facilita la comprensión del flujo de trabajo, desde la obtención de los datos hasta la generación de resultados analíticos útiles para la toma de decisiones institucionales.

Figura 1

Diagrama de Fases de Implementación



Fase 1 Limpieza y Normalización de los Datos

La primera fase del proyecto abarcó tanto la preparación técnica de los datos como su posterior depuración y normalización. Para iniciar este proceso, fue necesario establecer la conexión y configuración de la base de datos MySQL destinada al almacenamiento, consulta y gestión eficiente de la información extraída desde la red social X. Para ello se definieron los parámetros de acceso, la estructura de las tablas y los mecanismos de comunicación entre el sistema de extracción y el servidor, garantizando una transferencia de datos estable, confiable y segura. Esta infraestructura constituyó el soporte técnico fundamental sobre el cual se desarrollaron todas las etapas posteriores del análisis.

Dentro de la limpieza no se tuvo que realizar eliminación de registros duplicados al dataset, debido a que esta validación ya está contemplada dentro de la extracción propia a X.

Figura 2

Tabla Exploratoria de los Primeros Registros del Dataset de Tweets

First 10 Rows									
comentario	comment_date	usuario	palabras	clasificacion	puntuacion	fecha	mes	hora	longitud
@HaciendaBogota que entidad regula el precio de los arriendos...	2025-10-15T22:18:31	vivillena1006	entidad regula precio arriendos localidades Bogotá vías terr...	Negativo	-0.7	2025-10-15 22:18:36	2025-10	22	239
@consejodeestado @CNDJ_Col @Unilibrebog Sanciones por Abuso ...	2025-10-15T20:52:02	LouiseHCo	Sanciones Abuso Violacion derechos funcionarios embargos lo...	Negativo	-0.8	2025-10-15 20:52:04	2025-10	20	247
@Unaespecid @Bancolombia @HaciendaBogota @PoliciaColombia E...	2025-10-15T22:40:22	Jhonmauris	normal centros comerciales ofrecer planes turisticos citan L...	Negativo	-0.5	2025-10-15 22:40:28	2025-10	22	255
RT @JorgeGarzonAbg: @el_pais @CarlosFGalan Sentencias labora...	2025-10-16T04:22:00	JorgeGarzonAbg	Sentencias laborales cumplir Subredes Embargos	Neutral	-0.1	2025-10-16 04:22:04	2025-10	4	136
@el_pais @CarlosFGalan Sentencias laborales sin cumplir Subr...	2025-10-16T04:21:59	JorgeGarzonAbg	Sentencias laborales cumplir Subredes Embargos	Negativo	-0.7	2025-10-16 04:21:59	2025-10	4	116
@HaciendaBogota Facilitamos el pago de tus impuestos!!! Des...	2025-10-15T14:15:42	gabrielrios90	Descaro, ineficiencia, total media, hora, chat, interno, en...	Negativo	-0.8	2025-10-15 14:15:43	2025-10	14	265
@PGN_COL Freno ya a @HaciendaBogota Entidad Abusadora por qu...	2025-10-15T16:59:30	LouiseHCo	Freno HaciendaBogota Entidad Abusadora Realiza Embargos ILEG...	Negativo	-0.9	2025-10-15 16:59:30	2025-10	16	255
RT @chidalgo65: @lasilavacia @CarlosFGalan @MetroBogota @Jo...	2025-10-15T15:03:35	chidalgo65	nan	Neutral	0	2025-10-15 15:03:35	2025-10	15	142
RT @HaciendaBogota: Atención Estamos trabajando para mejorar...	2025-10-15T23:30:20	SanBibi2011	Atención trabajando mejorar servicios	Neutral	0	2025-10-15 23:30:20	2025-10	23	131
RT @JorgeGarzonAbg: @CarlosFGalan Esperamos	2025-10-16T00:41:57	JorgeGarzonAbg	Esperamos respuestas señor alcalde	Neutral	0	2025-10-16 00:41:57	2025-10	0	94

Una vez configurado el entorno de almacenamiento, se procedió a la extracción de datos mediante la API oficial de X, utilizando la librería Tweepy. Esta herramienta permitió recolectar tanto publicaciones recientes como históricas de la cuenta institucional de la Secretaría Distrital de Hacienda. Se obtuvo una muestra de 2.400 registros correspondientes al periodo comprendido entre el 9 de abril de 2025 y el 18 de octubre de 2025, lo cual aseguró un conjunto de datos amplio, representativo y adecuado para el análisis de percepción ciudadana. Como parte del proceso exploratorio, se generaron visualizaciones iniciales como curvas temporales de actividad y frecuencias por tipo de interacción que permitieron evaluar la coherencia y distribución temporal de la muestra recopilada.

En la **Figura 3** se puede observar el código utilizado en Python para la limpieza y normalización del campo texto en tweets, eliminando páginas, hashtags, menciones, caracteres especiales, números y emojis, y convirtiendo el texto a minúsculas.

Figura 3

Función de Limpieza de Texto para Tweets en Python

```
#####
# Funcion de limpieza del campo texto del tweet
def clean_text(text):
    text = re.sub(r'^RT[\s]+', '', text)
    text = re.sub(r'https?:\:\/\/.*[\r\n]*', '', text)#Elimina paginas
    text = re.sub(r'#', '', text)
    text = re.sub(r'@[A-Za-z0-9]+', '', text)
    text = text.lower() #pasa a minusculas
    text = re.sub("@[A-Za-z0-9_]+", "", text)
    text = re.sub("#[A-Za-z0-9_]+", "", text)
    text = re.sub(r"http\S+", "", text)
    text = re.sub(r"www.\S+", "", text)
    text = re.sub('(!?)', '', text)
    text = re.sub('\[.*?\]', '', text)
    text = re.sub("[^a-z0-9]", " ", text)
    text = re.sub(r"\[\S+\]", "", text) # Eliminar emoji
    return text
#####
```

Posteriormente, se dio inicio al proceso de limpieza y normalización del texto, una de las etapas más críticas para garantizar la calidad del análisis. Para ello se implementó en Python la función `clean_text()`, encargada de ejecutar transformaciones mediante expresiones regulares (regex) y reglas específicas de preprocesamiento. Esta función permitió eliminar enlaces web (URLs), menciones de usuario, hashtags, caracteres no relevantes, signos de puntuación, emojis, números aislados y palabras vacías (stopwords), así como otros elementos que podrían generar ruido en el análisis. Adicionalmente, el texto fue convertido a minúsculas para unificar criterios de comparación y evitar que las variaciones de capitalización afectaran la consistencia semántica.

Figura 4

Proceso de Limpieza y Normalización del Texto en el Análisis de Tweets



La fase de limpieza también contempló la corrección de espacios, la eliminación de saltos de línea innecesarios y la estandarización de expresiones propias del lenguaje digital, que con frecuencia aparecen abreviadas o distorsionadas. En los casos necesarios se aplicó la expansión de contracciones (por ejemplo, “q” → “que”), la eliminación de risas repetitivas como “jajajaja”

y la conversión de caracteres especiales mediante el estándar ASCII (American Standard Code for Information Interchange), un sistema de codificación que permite representar caracteres de forma uniforme para prevenir incompatibilidades entre plataformas y mejorar la interpretabilidad del texto.

Finalmente, se realizó un control de calidad del conjunto de datos procesado. Se descartaron mensajes duplicados, vacíos o sin contenido relevante, asegurando que cada registro aportara información significativa al análisis. Con ello, la fase dejó como resultado un conjunto de datos textual uniforme, depurado y semánticamente consistente, preparado para avanzar hacia la clasificación de sentimientos y la posterior visualización analítica.

En relación con la información contenida en el conjunto de datos, es pertinente aclarar que los nombres de usuario asociados a las publicaciones no serán anonimizados. Esto se justifica en que toda la información recolectada proviene de la red social X y corresponde a contenido completamente público, accesible libremente por cualquier usuario sin necesidad de autenticación ni permisos especiales. Según los lineamientos (*Datos abiertos*, s. f.) para el uso de datos en el sector público, los datos difundidos voluntariamente en plataformas abiertas pueden ser utilizados con fines académicos siempre que no se altere su naturaleza pública ni se haga un tratamiento que implique perfiles o decisiones automatizadas sobre individuos. Del mismo modo, la normativa colombiana (*Decreto 1377 de 2013 - Gestor Normativo*, s. f.) (Ley 1581 de 2012 y Decreto 1377 de 2013) establece que la protección de datos personales aplica principalmente a información privada o semiprivada; sin embargo, los datos de carácter público, como los nombres de usuario en redes sociales abiertas, no requieren autorización para su tratamiento cuando se emplean con fines de investigación. Bajo este marco, se mantiene la integridad de los datos tal como fueron publicados, dado que su inclusión permite comprender la

interacción entre ciudadanía y entidad pública sin vulnerar información sensible ni exponer datos privados.

Fase 2 Clasificación de Sentimientos

Dentro de la evaluación comparativa de los modelos de clasificación de sentimientos incluyendo arquitecturas basadas en transformers como BERT, RoBERTa y BERTO se optó por seleccionar TextBlob como modelo principal para el análisis. Esta decisión se fundamentó en criterios metodológicos, de transparencia y de eficiencia computacional, los cuales se detallan a continuación:

Metodología y Coherencia con el Enfoque del Proyecto

Para un estudio de carácter descriptivo, TextBlob ofrece un equilibrio adecuado entre simplicidad y precisión. Su enfoque léxico-estadístico, basado en técnicas clásicas de Procesamiento de Lenguaje Natural (PLN), permite obtener resultados consistentes sin incurrir en la complejidad estructural de los modelos avanzados. Esta característica lo convierte en una herramienta apropiada para proyectos cuyo objetivo principal es caracterizar tendencias generales más que alcanzar un nivel de clasificación altamente especializado.

Transparencia y Facilidad de Interpretación

A diferencia de los modelos basados en transformers, los cuales contienen millones de parámetros y operan como cajas negras difíciles de interpretar, TextBlob proporciona un proceso de clasificación más comprensible y trazable. Esto facilita explicar por qué un mensaje es clasificado como positivo, negativo o neutro, aspecto fundamental en investigaciones que requieren justificar y documentar claramente los criterios analíticos. En contraste, modelos como BERT o RoBERTa exigen análisis interpretativos avanzados y técnicas adicionales de explicabilidad (como LIME o SHAP), lo que incrementa la complejidad sin ser estrictamente necesario para los objetivos del estudio.

Eficiencia Computacional y Viabilidad Operativa

Desde la perspectiva del rendimiento, TextBlob presenta un procesamiento liviano y altamente eficiente, lo que reduce significativamente el consumo de memoria y los tiempos de ejecución. Esto es especialmente relevante dado que el volumen de datos puede crecer en futuros análisis. Mientras los modelos de transformers requieren hardware especializado (como GPU) y tiempos de inferencia considerablemente mayores, TextBlob mantiene un costo computacional bajo incluso a medida que aumenta la cantidad de publicaciones procesadas. Por tanto, garantiza sostenibilidad técnica a largo plazo.

Figura 5

Análisis de Sentimientos de Tweets en la Red Social X con TextBlob



La aplicación del análisis de sentimientos no solo permitió clasificar los mensajes según su orientación emocional, sino que también brindó una estructura interpretativa útil para comprender cómo los ciudadanos perciben a la Secretaría Distrital de Hacienda en distintos

contextos comunicativos. Este enfoque facilitó la identificación de tendencias, reacciones ante temas específicos y posibles focos de inconformidad o aprobación dentro de la conversación digital.

En cuanto al rendimiento del modelo, las métricas globales indicaron que TextBlob alcanzó una exactitud del 75%, lo cual resulta aceptable para una primera aproximación basada en un método léxico. Sin embargo, un examen detallado de las métricas por clase (precisión, recall y F1-score) permitió evidenciar un comportamiento desigual entre las categorías de sentimiento. La clase neutra obtuvo el mejor desempeño, lo cual es coherente con la naturaleza del modelo, que tiende a identificar con mayor facilidad textos con baja carga emocional. Además, es importante resaltar que el conjunto de datos incluye no solo los mensajes publicados por la ciudadanía, sino también las respuestas emitidas por los funcionarios de la entidad. Estas interacciones fueron intencionalmente conservadas en el análisis, ya que permiten evaluar el tono y el tipo de sentimiento con el cual la institución responde a las inquietudes y comentarios de los usuarios. La inclusión de estos mensajes institucionales explica, en parte, la alta proporción de clasificaciones neutras observadas en los resultados, dado que las respuestas oficiales suelen emplear un lenguaje formal, informativo y carente de carga emocional.

Por el contrario, las clases positiva y negativa presentaron puntajes considerablemente más bajos. No obstante, este resultado no debe interpretarse únicamente como una limitación de la librería, sino como una consecuencia directa de la distribución desequilibrada del conjunto de datos. En la muestra analizada se encontró un número muy reducido de mensajes positivos y negativos, mientras que la gran mayoría correspondía a textos neutros. En este tipo de escenarios, incluso un pequeño número de errores tiene un impacto proporcional mucho mayor en las métricas, incrementando los falsos positivos y falsos negativos. Esta desproporción,

sumada al uso frecuente de lenguaje informal, sarcasmo, abreviaturas y expresiones locales de Bogotá y Colombia, dificultó aún más la correcta identificación del tono emocional.

Dado este panorama y con el fin de mejorar la calidad del análisis, se procedió a realizar una clasificación manual complementaria del conjunto de datos. Para esta tarea se definieron criterios claros que permitieran asignar de manera consistente cada mensaje a una de las tres categorías emocionales:

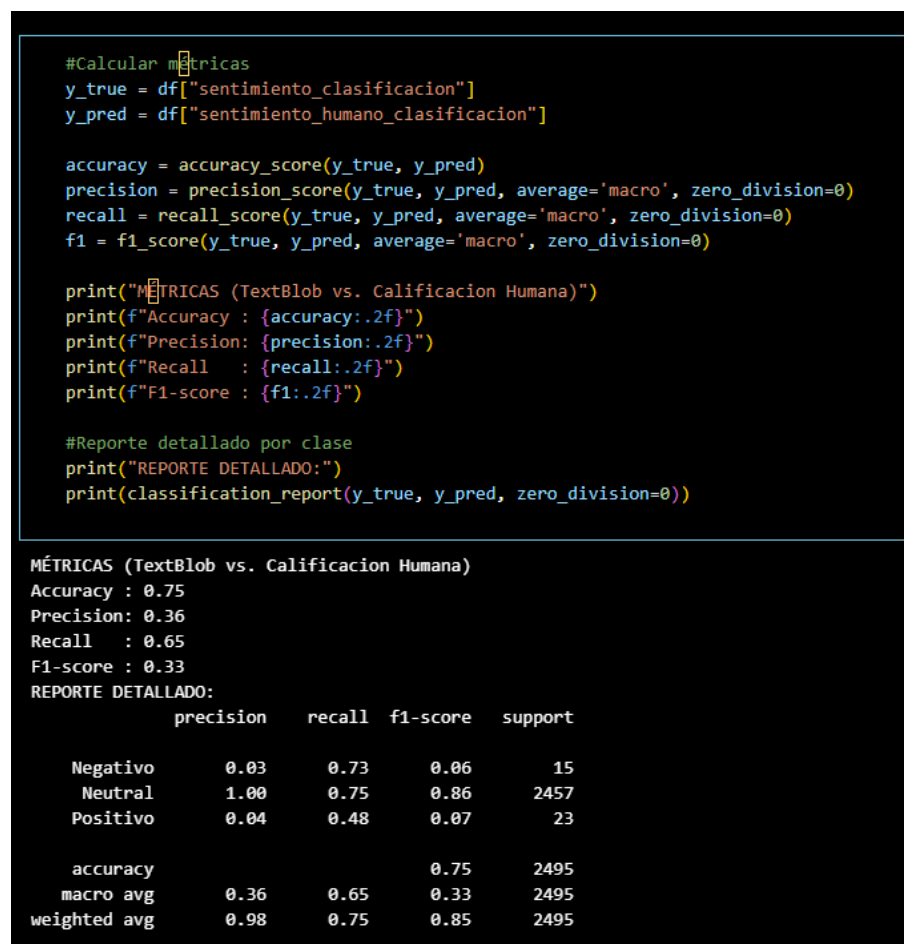
Sentimiento positivo: mensajes que expresaban aprobación, felicitaciones, satisfacción, agradecimiento o valoración favorable hacia la entidad o sus acciones.

Sentimiento negativo: publicaciones que incluían quejas, inconformidades, reclamos, críticas o percepciones desfavorables, ya fuera por trámites, atención o decisiones institucionales.

Sentimiento neutro: textos informativos, consultas, anuncios, menciones sin connotación emocional o mensajes cuyo contenido no permitía inferir una postura clara del autor.

Esta clasificación manual permitió validar el comportamiento del modelo, equilibrar mejor el análisis y fortalecer la interpretación final de los resultados. En términos generales, esta fase aportó una comprensión cuantitativa y cualitativa de la percepción ciudadana frente a la entidad, al tiempo que evidenció la necesidad de explorar modelos más robustos, como enfoques supervisados o redes neuronales basadas en transformadores, para mejorar el rendimiento en futuras iteraciones del proyecto.

En la **Figura 6** podemos ver el cálculo e impresión de métricas de desempeño (accuracy, precision, recall y F1-score) para la clasificación de sentimientos en tweets, comparando la clasificación automática por TextBlob frente a la calificación humana.

Figura 6*Evaluación de Métricas de Clasificación de Sentimientos*

Como se observa en la **Figura 7**, el mapa de calor permite visualizar de manera precisa el rendimiento del modelo según cada categoría de sentimiento. La clase Neutro se destaca como la mejor clasificada, evidenciando una precisión alta y un F1-score superior en comparación con las demás clases. Esto confirma que el modelo tiende a interpretar la mayoría de los mensajes como neutrales y que logra hacerlo con un nivel aceptable de consistencia.

En contraste, las categorías Negativo y Positivo presentan un desempeño considerablemente inferior, reflejado en F1-scores bajos y valores de precisión reducidos.

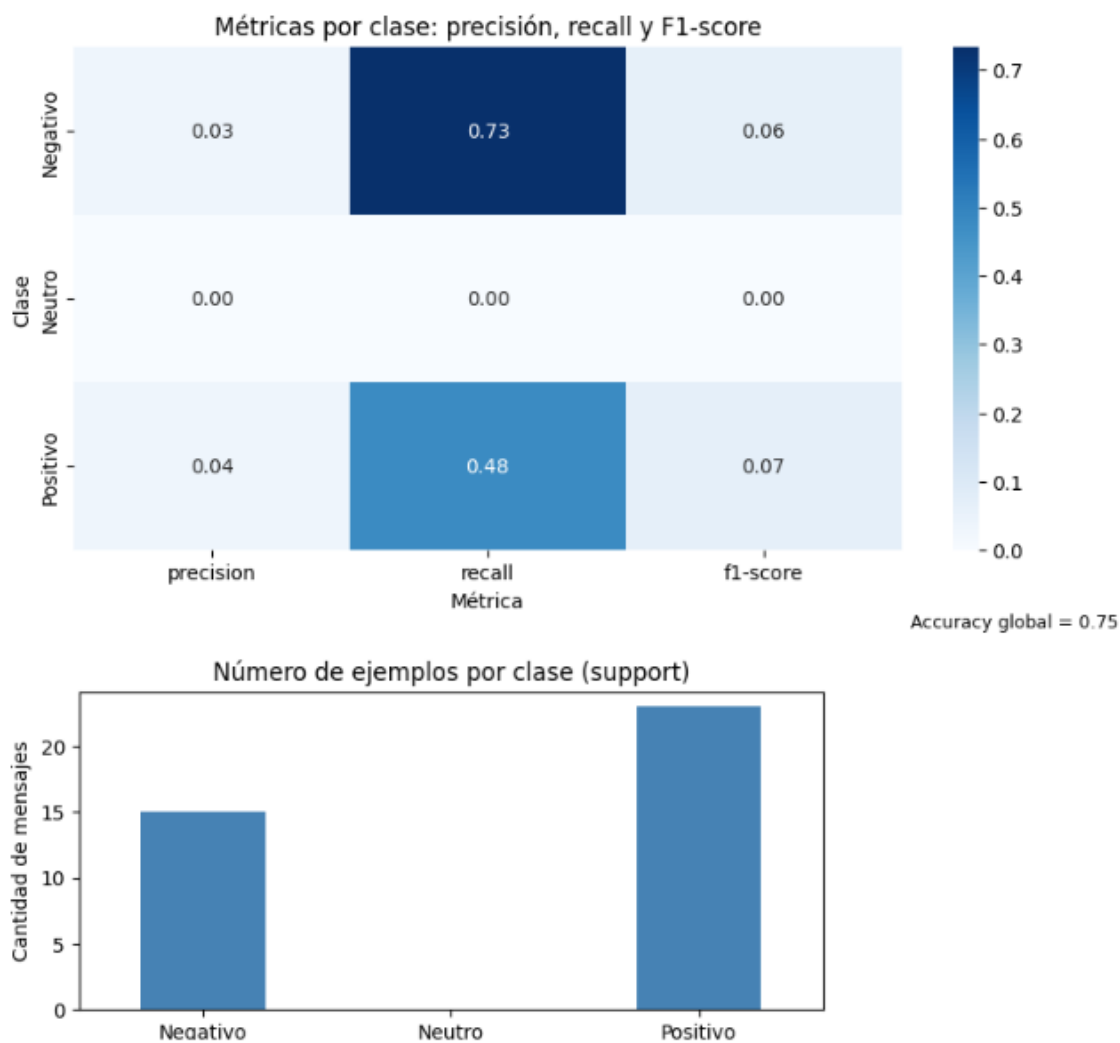
Aunque la clase Negativo muestra un recall elevado lo que indica que el modelo identifica una parte importante de los mensajes realmente negativos su baja precisión demuestra que asigna esta etiqueta incorrectamente en múltiples ocasiones. Esta disparidad evidencia que TextBlob enfrenta dificultades para capturar matices emocionales específicos, especialmente cuando el lenguaje incluye ironía, sarcasmo, abreviaciones o expresiones coloquiales frecuentes en la interacción ciudadana en redes sociales.

Es importante mencionar que estos resultados también están influenciados por el desbalance del conjunto de datos, donde los mensajes con sentimiento positivo y negativo representan una proporción muy reducida frente a los neutros. En conjuntos pequeños, como el utilizado en este proyecto, incluso unos pocos errores generan variaciones porcentuales grandes en las métricas de estas clases. Por ello, parte de las limitaciones observadas no provienen únicamente del modelo, sino también del tamaño y distribución del dataset disponible.

No obstante, a medida que el proyecto evolucione y sea posible incorporar un volumen mayor de datos y anotaciones manuales, se espera que el rendimiento para las categorías Positivo y Negativo mejore significativamente. Esto permitiría capturar patrones emocionales más complejos y aportar una visión más fina del comportamiento ciudadano frente a la gestión de la Secretaría Distrital de Hacienda.

Figura 7

Mapa de Calor con Métricas de Desempeño por Clase en Clasificación de Sentimientos



Fase 3 Análisis Temático y Visualización de Resultados

En la última fase del proyecto se integraron todos los resultados obtenidos en las etapas previas dentro de un tablero de control interactivo, diseñado para facilitar la comprensión del comportamiento ciudadano en la red social X. Este dashboard permitió organizar la información procesada en visualizaciones intuitivas, transformando los datos textuales y las métricas de análisis en indicadores claros y accesibles para la toma de decisiones institucionales. La herramienta fue construida con el propósito de ofrecer una lectura dinámica de los hallazgos, permitiendo al usuario final filtrar, explorar y contrastar la información según distintos criterios de interés.

Para la identificación de temas predominantes dentro del conjunto de mensajes analizados, se empleó un conjunto de técnicas de minería de texto que permitieron explorar la estructura semántica de los datos desde diferentes niveles. En primera instancia, se realizó un análisis de frecuencia de palabras, mediante el cual se identificaron los términos más recurrentes en los mensajes. Esta técnica facilitó la detección de conceptos clave asociados a inquietudes ciudadanas, ya que las palabras con mayor frecuencia suelen representar los focos temáticos más discutidos.

Posteriormente, se aplicó la matriz de términos y documentos (TF-IDF), un método que pondera la importancia de cada palabra dentro del conjunto total de mensajes al considerar su frecuencia relativa. A diferencia del conteo simple, TF-IDF permitió distinguir términos verdaderamente relevantes aquellos que aportan significado específico de palabras comunes que, aunque frecuentes, no necesariamente contribuyen a la identificación de temas. Este enfoque hizo posible resaltar vocabularios característicos de conversaciones particulares, mejorando la capacidad del análisis para detectar tópicos emergentes.

Asimismo, se utilizó la agrupación semántica mediante n-gramas, técnica que permite analizar combinaciones de palabras consecutivas (por ejemplo, bigramas y trigramas). Esto resultó fundamental para capturar expresiones compuestas como “pago impuesto”, “atención ciudadanía” o “plazos tributarios”, las cuales no pueden identificarse mediante análisis de palabras aisladas. El estudio de n-gramas contribuyó, por tanto, a reconocer relaciones sintácticas y semánticas que muestran cómo los usuarios estructuran sus opiniones.

De manera complementaria, se incorporaron técnicas exploratorias como las nubes de palabras, la coocurrencia de términos y el conteo detallado de bigramas y trigramas, herramientas que facilitaron la visualización intuitiva de los conceptos más mencionados y de su relación con otros términos relevantes. Estas representaciones gráficas permitieron observar patrones de conversación, identificar grupos de palabras asociadas a sentimientos específicos y reconocer temas que concentraron mayor volumen de discusión entre los usuarios.

En conjunto, estas técnicas ofrecieron una visión integral de la dinámica temática de los mensajes analizados, permitiendo no solo identificar los tópicos centrales, sino también comprender la forma en que fueron construidas las narrativas ciudadanas dentro de la red social X.

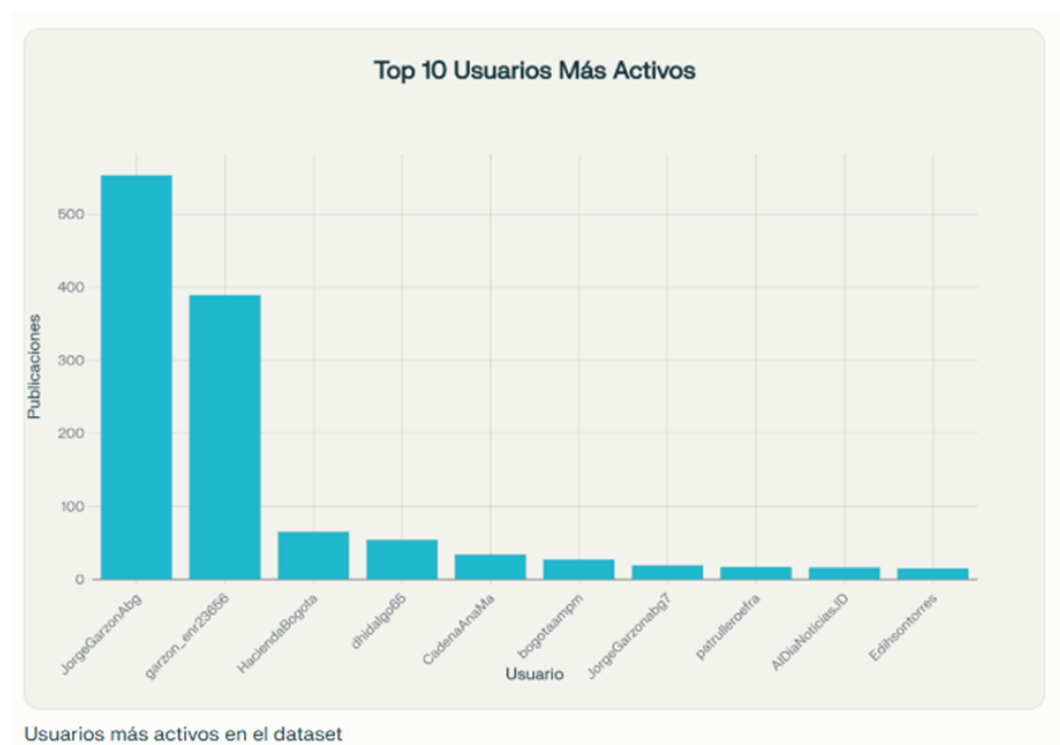
El dashboard consolidado ofreció una perspectiva integral del comportamiento de los usuarios en la plataforma, incorporando diversas visualizaciones complementarias. La nube de palabras resaltó los términos más repetidos en los mensajes, reflejando los conceptos y temas que captaron mayor interés entre la ciudadanía. Por su parte, los gráficos de frecuencia y líneas temporales permitieron observar la evolución del sentimiento a lo largo de los días, evidenciando picos de mensajes negativos o positivos asociados a eventos, noticias o publicaciones

institucionales. Esta información temporal aportó un contexto clave para interpretar reacciones frente a decisiones administrativas, cambios normativos o anuncios relevantes.

Asimismo, el panel de usuarios más activos proporcionó información valiosa sobre quiénes participan con mayor recurrencia en la conversación digital, lo que permitió identificar perfiles influyentes o cuentas que concentran volúmenes significativos de interacción. El indicador de polaridad, complementado con métricas agregadas, ofreció una visión clara del tono general de la conversación, facilitando la identificación de cambios en el sentimiento predominante y posibles señales de alerta comunicacional. Finalmente, la tabla detallada de mensajes que incluyó fecha, texto original, clasificación emocional y polaridad numérica permitió profundizar en la revisión cualitativa de los datos, asegurando la transparencia y trazabilidad del proceso analítico.

Figura 8

Top 10 de Usuarios Más Activos en la Conversación Digital Sobre la SDH



En conjunto, esta fase final aseguró una visualización clara, completa y operativa de toda la información procesada, convirtiendo el análisis de datos en una herramienta estratégica para la entidad. El tablero de control va a fortalecer la capacidad de la Secretaría Distrital de Hacienda para monitorear la percepción ciudadana, y también fortalecer la formulación de estrategias comunicacionales basadas en evidencia real y actualizada. Con esto, el proyecto demuestra cómo el análisis de datos y la visualización avanzada pueden apoyar la toma de decisiones públicas y mejorar la interacción entre ciudadanía e institución.

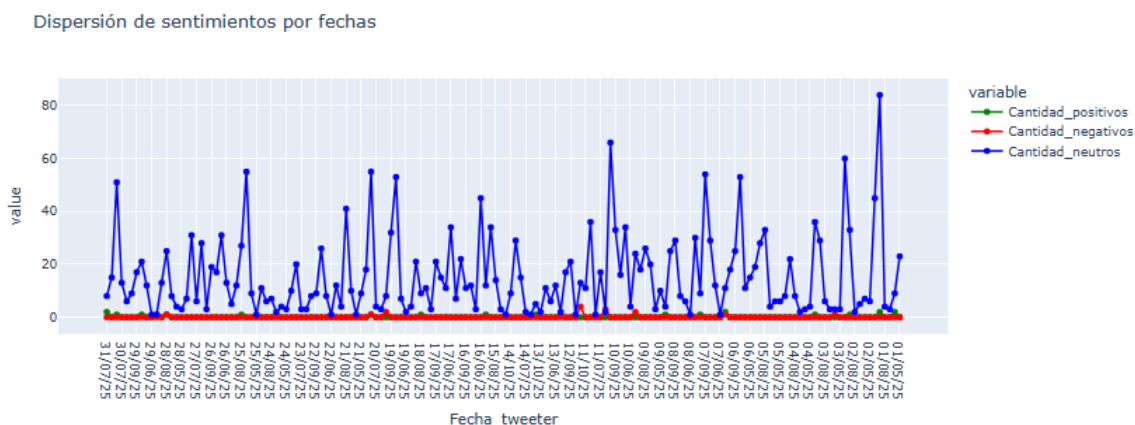
Como podemos ver en la **Figura 9** la nube de palabras evidencia que la conversación ciudadana en la red social X se concentra en torno a temas como “sentencias laborales”, “trabajadores”, “impuesto”, “pago”, “recaudo”, “embargo” y “nuevo contralor”, lo que sugiere que las principales preocupaciones giran alrededor del cumplimiento de obligaciones laborales, la gestión del impuesto predial y los procesos de cobro. La alta presencia de términos asociados a “Bogotá”, “secretaría”, “proyecto” e “inversión” indica que los usuarios relacionan directamente a la entidad con decisiones de ciudad y con la forma en que se comunican cambios normativos o tributarios. Para la Secretaría Distrital de Hacienda, estos resultados son valiosos porque permiten priorizar mensajes aclaratorios y campañas pedagógicas sobre los temas que más se repiten, así como diseñar estrategias de comunicación específicas frente a las problemáticas que generan mayor volumen de conversación y posibles riesgos reputacionales.

En la **Figura 10** se observa cómo se comparan, a lo largo del período analizado, las frecuencias diarias de mensajes clasificados como positivos, negativos y neutros. Se observa que la categoría predominante es la de mensajes neutros, los cuales presentan una variabilidad considerable con picos recurrentes de actividad. Por el contrario, los mensajes positivos y negativos aparecen en volúmenes significativamente menores, lo que sugiere que la conversación digital sobre la Secretaría Distrital de Hacienda tiende a mantenerse informativa o descriptiva más que emocional.

Este comportamiento también explica, en parte, las diferencias en el rendimiento del modelo de análisis de sentimientos, dado que la baja representación de mensajes con polaridad marcada puede generar mayor sensibilidad a errores en dichas clases. En conjunto, el gráfico permite identificar momentos específicos en los que aumentó la interacción ciudadana y, potencialmente, evaluar la relación entre estas variaciones y eventos institucionales, comunicados oficiales o situaciones coyunturales.

Figura 10

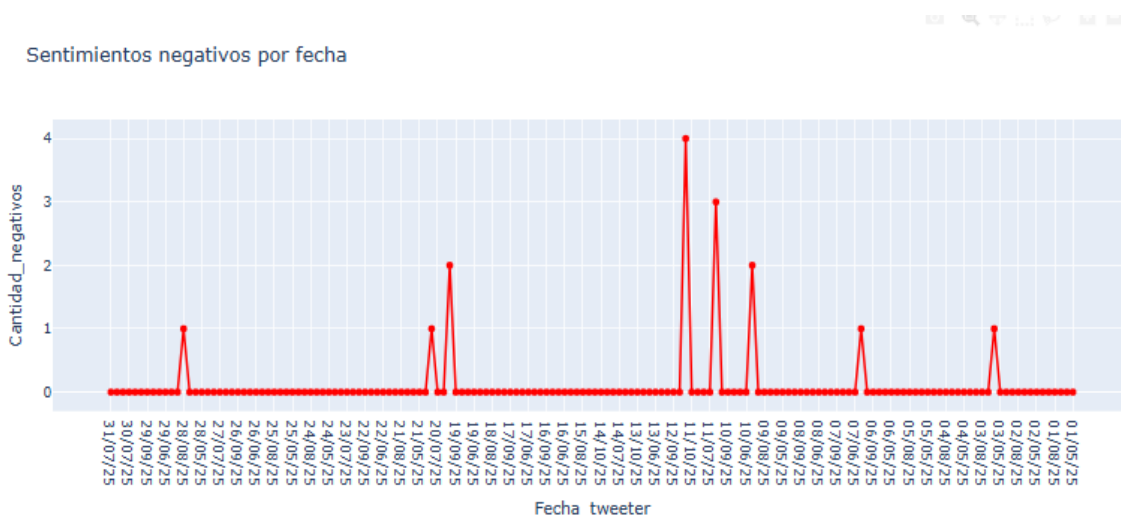
Dispersión Temporal de Sentimientos en Tweets



La evolución temporal de sentimientos negativos en la red social X permite identificar los días en los que se concentró un mayor volumen de mensajes críticos hacia la Secretaría Distrital de Hacienda, evidenciando picos específicos de inconformidad ciudadana. Estos incrementos puntuales en la cantidad de tweets negativos se asocian con eventos, anuncios o gestiones institucionales particulares, por lo que constituyen señales de alerta que orientan a la entidad sobre momentos en los que fue percibida con mayor descontento. En el marco del proyecto, este análisis temporal complementa la clasificación de sentimientos y las visualizaciones del tablero de control, aportando un insumo clave para priorizar la revisión de fechas críticas y diseñar acciones de comunicación o mejora del servicio basadas en la evidencia observada en la conversación digital.

Figura 11

Evolución Temporal de Sentimientos Negativos en la Red Social X



En la **Figura 12** se puede observar los usuarios con mayor actividad en la red social X evidencia qué cuentas concentran el volumen más alto de publicaciones en la muestra analizada, destacando especialmente perfiles como JorgeGarzonAbg, garzon_enr23656 y la cuenta

institucional HaciendaBogota, que acumulan varios cientos de interacciones. Esta distribución permite identificar actores particularmente influyentes en la conversación digital sobre la Secretaría Distrital de Hacienda, ya sea por su rol institucional, mediático o de liderazgo de opinión. En el marco del proyecto, esta información resulta clave para contextualizar los resultados del análisis de sentimientos, ya que ayuda a diferenciar mensajes provenientes de usuarios altamente activos de aquellos emitidos por ciudadanos ocasionales, lo que aporta matices a la interpretación de la percepción ciudadana en X.

Figura 12

Usuarios con Mayor Actividad en X Según Conteo de Publicaciones



↕conteo	↕usuario_twitteador
filter data...	
553	JorgeGarzonAbg
389	garzon_enr23656
65	HaciendaBogota
54	dhidalgo65
34	CadenaAnaMa
27	bogotaampn
19	JorgeGarzonabg7
17	patrulleroefra
16	ALDiaNoticiasJD
15	Edihsontorres

La **Figura 13** nos permite ver el conteo diario de tweets realizados en la red social X de la Secretaría Distrital de Hacienda, también muestra la variación en el número de publicaciones

registradas por fecha, permitiendo identificar jornadas con actividad excepcionalmente alta frente a días de baja conversación. Estos picos de mensajes suelen corresponder a eventos específicos, anuncios institucionales o controversias puntuales que detonaron mayor interacción ciudadana en la red social X. En el contexto del proyecto, este indicador temporal complementa el análisis de sentimientos al ofrecer una lectura cuantitativa del volumen de participación, lo que facilita relacionar cambios en la intensidad del debate digital con la aparición de temas sensibles o decisiones de la entidad que generaron mayor atención pública.

Figura 13

Conteo Diario de Tweets Sobre la Secretaría Distrital de Hacienda



↕fecha	↕conteo
filter data...	
31/07/25	10
30/09/25	15
30/08/25	52
30/07/25	13
30/05/25	6
30/04/25	9
29/09/25	17
29/08/25	22
29/07/25	12
29/06/25	1

Como se puede observar en la **Figura 14** la cual corresponde a un gráfico de distribución porcentual de mensajes por polaridad en la red social X y se evidencia un claro predominio de

publicaciones clasificadas como neutrales, que representan aproximadamente el 98,5% del total, frente a proporciones muy reducidas de mensajes positivos y negativos. Este resultado confirma lo observado en la evaluación del modelo de TextBlob, donde la clase neutra obtuvo el mejor desempeño y tendió a agrupar buena parte de los mensajes con baja carga emocional. En el marco del proyecto, esta distribución refuerza la necesidad de interpretar los hallazgos de manera crítica, considerando que la naturaleza léxica del modelo y el lenguaje informal de las redes pueden sesgar la clasificación hacia la neutralidad, al tiempo que ofrece una primera aproximación cuantitativa al tono general de la conversación ciudadana sobre la Secretaría Distrital de Hacienda.

Figura 14

Distribución Porcentual de Mensajes por Polaridad en X



En la **Figura 15** se observa el detalle de los tweets analizados para cada registro, la fecha del mensaje, el texto original publicado en la red social X, el usuario emisor y el resultado de la evaluación de sentimiento asignado por el modelo, así como la fecha de registro en la base de datos. Esta visualización permite rastrear de manera transparente cómo se clasificó cada publicación, facilitando la revisión cualitativa de casos específicos y la validación de los resultados cuantitativos obtenidos en el análisis de sentimientos. En el marco del proyecto, esta vista tabular actúa como capa de trazabilidad del proceso analítico, ya que conecta las métricas agregadas del dashboard con los mensajes ciudadanos individuales que les dan origen, fortaleciendo la confiabilidad e interpretabilidad de los hallazgos.

Figura 15

Detalle Tabular de Tweets Analizados en Estudio de Sentimientos

DETALLE DE LOS TWEETS

Fecha_tweet	Texto_tweet	Usuario_tweetador	Usuario_tweetado	Eval_Sentimiento_resul	Fecha_registro
9/04/2025	@HaciendaBogota Ya lo hice A A, EMM-	nancynxtr		Neutral	2025-04-09T00:00:00
9/04/2025	@HaciendaBogota al pagar el predial me sale un error de 'cáma'digo de recaudo no válidlo'.	SantiagoBalcerero		Neutral	2025-04-09T00:00:00
9/04/2025	@HaciendaBogota por qué si el impuesto vehicular para carros con valor menor a 56 millones es el 1.5% hacienda BogotA estA liquidando ni impuesto con el 1.7%? Me gustarA conocer su respuesta.	PABLODorado765		Neutral	2025-04-09T00:00:00
9/04/2025	@HaciendaBogota Simplemente pido una explicaciAn y ustedes la pueden dar abierta al pAblico les parece?	PABLODorado765		Neutral	2025-04-09T00:00:00
9/04/2025	@CarlosFGalan @GalanAlcaldeBog @SectorSalud @paulabolivar Alcalde @CarlosFGalan: mientras se prioriza la persecuciAn de delitos en tiempo real mAjs de 500 trabajadores de la salud con sentencias judiciales firmes continAn sin recibir el			Positivo	2025-04-09T00:00:00
13/04/2025	@HaciendaBogota Hola buenos dAs como todos los aAs es una pesadilla pagar el impuesto predial. Ya estAn pensando en correr la fecha?? https://t.co/1cozTb1j	rafael_valero		Neutral	2025-04-13T00:00:00
13/04/2025	@HaciendaBogota Ya envié el caso espero respuesta	rafael_valero		Neutral	2025-04-13T00:00:00
13/04/2025	@BancoAgrario @CristoBustos @jsanchezcristo Es importante seAtor Alcalde el cumplimiento de sentencias laborales ejecutoriadas con eso el distrito acata las decisiones judiciales que es un deber legal y no debe tener condicionamientos. @CarlosFGalan @Maci	JorgeGarzonAbg		Positivo	2025-04-13T00:00:00
14/04/2025	RT @JorgeGarzonAbg: @BancoAgrario @CristoBustos @jsanchezcristo Es importante seAtor Alcalde el cumplimiento de sentencias laborales ejecutoriadas	JorgeGarzonAbg		Neutral	2025-04-14T00:00:00
14/04/2025	@HaciendaBogota @CadenaAnaMa @Citytv @RadioColombia dudas no han podido ser aclaradas ni el supercade 30 ni el punto de la 114 ni el supercade 20 de julio. Me sugieren escribir un correo. Sin embargo considero que ello no es lo ideal dado la especAfico	MichelRiff		Positivo	2025-04-14T00:00:00
14/04/2025	RT @HaciendaBogota: A A, A A TenEnCuenta El vienes 25 de abril vence el plazo para pagar el #PredialBogotA 2025 A A, A A con descuento del 10 %. A A, A A SigúfA!	SanBibi2011		Neutral	2025-04-14T00:00:00

Conclusiones

En conjunto, las tres fases desarrolladas reflejan de manera integral la aplicación de los conocimientos teóricos y prácticos adquiridos en la Especialización en Ciencia de Datos y Analítica, evidenciando la capacidad de trasladar los fundamentos de la analítica avanzada a un contexto real del sector público. Este proyecto no solo representa un ejercicio técnico de programación y análisis de datos, sino también una propuesta estratégica orientada a fortalecer la gestión institucional y la transparencia administrativa mediante el uso inteligente de la información generada por los ciudadanos en entornos digitales.

La implementación de técnicas de minería de texto, procesamiento de lenguaje natural (PLN) y análisis de sentimientos, complementadas con procesos de depuración, categorización y visualización de la información, permitió estructurar un flujo de trabajo robusto que convierte datos no estructurados en conocimiento accionable y medible. Este enfoque contribuye directamente al fortalecimiento institucional de la Secretaría Distrital de Hacienda, al ofrecer una herramienta tecnológica capaz de monitorear de forma sistemática la percepción ciudadana, detectar patrones emocionales y temáticos, y anticipar posibles áreas de mejora o riesgo reputacional.

Asimismo, el proyecto consolida una metodología replicable para otras entidades públicas interesadas en incorporar la analítica de datos en sus procesos de toma de decisiones. Su valor radica en demostrar que, mediante el aprovechamiento de los datos sociales y el uso de la inteligencia artificial, es posible optimizar los canales de comunicación institucional, diseñar estrategias más efectivas y promover una gestión pública basada en evidencia. En suma, este trabajo refleja el potencial transformador de la ciencia de datos como instrumento para fortalecer

la relación entre el Estado y la ciudadanía, impulsando una administración más participativa, transparente y orientada al servicio ciudadano.

Recomendaciones

Implementar la solución propuesta de forma oficial en la Secretaría Distrital de Hacienda, integrándola a los sistemas existentes de gestión y comunicación institucional.

Monitorear continuamente la percepción ciudadana, manteniendo actualizado el sistema de recolección y análisis de datos en redes sociales, con énfasis en temas sensibles o de alto impacto público.

Ampliar el análisis a los diferentes canales de comunicación que cuenta la entidad (redes sociales, encuestas, medios digitales) y exploren nuevas técnicas de inteligencia artificial para mejorar la calidad del análisis emocional y temático.

Fortalecer las capacidades del personal institucional mediante capacitaciones en analítica de datos y gestión de información pública, garantizando la sostenibilidad y el aprovechamiento continuo de la herramienta en los procesos de toma de decisiones.

Establecer indicadores de desempeño y reportes periódicos que permitan evaluar el impacto de la herramienta en la mejora de la comunicación institucional y en la percepción ciudadana, promoviendo una cultura de gestión basada en datos dentro de la Secretaría Distrital de Hacienda.

Referencias

- Alexander Pak & Patrick Paroubek. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. En N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC`10)*. European Language Resources Association (ELRA). <https://aclanthology.org/L10-1263/>
- análisis de sentimientos en redes sociales—Resultados de la búsqueda—EBSCO Discovery Service.* (s. f.). Recuperado 30 de marzo de 2025, de <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/c/qcagk4/search/results?q=análisis%20de%20sentimientos%20en%20redes%20sociales&autocorrect=y&expanders=fullText&expanders=concept&limiters=FT1%3AY&resetPageNumber=true&searchMode=all&searchSegment=all-results>
- Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review. *SN Computer Science*, 3(1), 74. <https://doi.org/10.1007/s42979-021-00958-1>
- Carangui, A. K. U., Almeida, S. C., Matamoros, R. P., & Morocho, G. (2024). El análisis de los sentimientos con Inteligencia Artificial como estrategia de las relaciones públicas. *Ciencia Latina Revista Científica Multidisciplinar*, 8(5), Article 5. https://doi.org/10.37811/cl_rcm.v8i5.14174
- Carlos Busón Buesa. (2020). La minería de opinión para el análisis del discurso de odio en las redes sociales: Opinion mining for the analysis of hate speech in social networks. *Commons: revista de comunicación y ciudadanía digital*, 9(1), 119-159. <https://doi.org/10.25267/commons.2020.v9.i1.5>

Datos abiertos. (s. f.). Recuperado 8 de diciembre de 2025, de

<https://gobiernodigital.mintic.gov.co/portal/Iniciativas/Datos-abiertos/>

Decreto 1377 de 2013—Gestor Normativo. (s. f.). Recuperado 8 de diciembre de 2025, de

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=53646>

García, L. M. (2024). *Análisis de sentimientos en Instagram usando ChatGPT.*

<https://uvadoc.uva.es/handle/10324/71261>

García Merayo, M. D. L. M., Requena Román, D., Cerro del Pino, E., & de Diego Berlanga, Á.

(2023). *Análisis de sentimientos en redes sociales.*

<https://research.ebsco.com/linkprocessor/plink?id=e12cb01e-887d-3285-8854-a59f98407925>

Lovera, F. A., & Cardinale, Y. (2023). Análisis de sentimientos en Twitter: Un estudio

comparativo. *Revista Científica de Sistemas e Informática*, 3(1).

https://portal.amelica.org/ameli/journal/535/5354040002/html/?utm_source=chatgpt.com

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand

sentiments. *Expert Systems with Applications*, 40(10), 4241-4251.

<https://doi.org/10.1016/j.eswa.2013.01.019>

Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A., & Wilson, T. (2013). SemEval-

2013 Task 2: Sentiment Analysis in Twitter. En S. Manandhar & D. Yuret (Eds.), *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2:*

Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) (pp. 312-320). Association for Computational Linguistics.

<https://aclanthology.org/S13-2052/>

- Reyes, N. S., & Moreno, G. E. T. (2024). Análisis de sentimientos en datos de redes sociales: Aplicación de técnicas de procesamiento de lenguaje natural y machine learning para analizar opiniones y sentimientos en datos de redes sociales en el contexto de sistemas de información. *Dominio de las Ciencias*, 10(1), Article 1.
<https://doi.org/10.23857/dc.v10i1.3714>
- TFG_MARIA_BELEN_ALARCON_RAMOS.pdf*. (s. f.). Recuperado 28 de marzo de 2025, de https://oa.upm.es/70960/1/TFG_MARIA_BELEN_ALARCON_RAMOS.pdf?utm_source=chatgpt.com
- Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1), 178-185.
<https://doi.org/10.1609/icwsm.v4i1.14009>
- Valencia, R. A. M., Calle, O. R., Carrasco, G. E. B., & Ordinola, V. H. R. (2024). Modelo con Minería de Texto y Diccionarios de Sentimientos en Redes Sociales: Transporte público América Latina. *RISTI: Revista Ibérica de Sistemas e Tecnologias de Informação*, Extra 73, 133-150.
- Yenkikar, A., Babu, C. N., & Hemanth, D. J. (2022). SENTINET: A deep sentiment analysis network for political media bias detection. *DYNA - Ingeniería e Industria*, 97(6), 645-651. <https://doi.org/10.6036/10593>