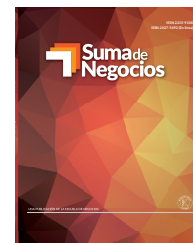




SUMA DE NEGOCIOS



Artículo de investigación

Estimación de la incertidumbre en Ecopetrol: aplicación de PNL y modelos Bert para la caracterización empresarial

Fernando Cárdenas García¹, Gregorio Izquierdo Llanes² y José Gabriel Orosco Jiménez³

¹ Magíster en Economía. Profesor, Universidad Militar Nueva Granada, Bogotá, Colombia (autor de correspondencia). Correo electrónico: jose.cardenas@unimilitar.edu.co

² Doctor en Ciencias Económicas y Empresariales. Profesor, Universidad Nacional de Educación a Distancia, Madrid, España. Correo electrónico: gizquierdo@cee.uned.es

³ Estudiante de Economía, Universidad Sergio Arboleda, Bogotá, Colombia. Correo electrónico: j.gabrielorosco@gmail.com

INFORMACIÓN DEL ARTÍCULO

Recibido el 29 de diciembre de 2024

Aceptado el 20 de julio de 2025

Online el 22 de septiembre de 2025

JEL Codes:

G14, G41, C45, Q40

Palabras clave:

Incertidumbre,
procesamiento de lenguaje natural,
modelos BERT,
análisis de sentimientos,
volatilidad financiera,
Ecopetrol.

Keywords:

Uncertainty,
natural language processing,
BERT,
sentiment analysis,
financial volatility,
Ecopetrol.

RESUMEN

Introducción/objetivo: este estudio busca desarrollar un indicador de incertidumbre para Ecopetrol, una de las principales empresas del sector energético en Colombia, mediante el análisis de sentimientos en noticias y reportes financieros. La investigación aborda un vacío en la literatura al combinar análisis de sentimientos con datos específicos del sector energético.

Metodología: mediante técnicas avanzadas de procesamiento de lenguaje natural (PLN) y modelos de aprendizaje profundo como BERT, se busca cuantificar de qué manera la percepción del mercado, reflejada en textos, influye en la volatilidad de los activos de la compañía. Se emplearon modelos de lenguaje como FinBERT, XLM-RoBERTa, BETO y DeBERTa, para analizar sentimientos en textos multilingües (español e inglés).

Resultados: los picos de entropía en titulares (español/inglés) coincidieron con eventos clave, mostrando una correlación con la volatilidad de Ecopetrol; mientras que los análisis con FinBERT y XLM-RoBERTa, se destacan en la precisión para textos financieros y la captura de percepciones multilingües.

Conclusiones: se identifica una relación negativa entre sentimiento positivo y volatilidad, sugiriendo que el optimismo reduce la incertidumbre del mercado. El estudio demuestra que el análisis de sentimientos con IA es una herramienta valiosa para predecir volatilidad en mercados emergentes.

Estimating uncertainty at Ecopetrol: Applying NLP and BERT models for business characterization

ABSTRACT

Introduction/Objective: This study aims to develop an uncertainty indicator for Ecopetrol, one of the leading companies in Colombia's energy sector, through sentiment analysis of news articles and financial reports. The research addresses a gap in the literature by combining sentiment analysis with energy sector-specific data.

Methodology: Using advanced Natural Language Processing (NLP) techniques and deep learning models such as BERT, the study seeks to quantify how market perception

—reflected in textual data—affects the volatility of the company's assets. Language models such as FinBERT, XLM-RoBERTa, BETO, and DeBERTa were used to analyze sentiment in multilingual texts (Spanish and English).

Results: Spikes in headline entropy (Spanish/English) aligned with key events, showing a correlation with Ecopetrol's volatility. Sentiment analyses using FinBERT and XLM-RoBERTa stood out for their accuracy in financial texts and their ability to capture multilingual perceptions.

Conclusions: A negative relationship was identified between positive sentiment and volatility, suggesting that optimism reduces market uncertainty. The study demonstrates that AI-based sentiment analysis is a valuable tool for forecasting volatility in emerging markets.

Introducción

En un contexto financiero cada vez más dinámico y complejo, los textos de noticias, los reportes financieros y sus interpretaciones inciden sobre la incertidumbre y se han convertido en un elemento crucial que afecta las decisiones de inversión y la estabilidad de los mercados, así como la gestión reputacional. La volatilidad de los activos, en particular, está influenciada por la forma en que los participantes del mercado perciben y reaccionan ante la información disponible. En este sentido, el análisis de noticias y reportes financieros ha demostrado ser una fuente valiosa de datos que puede proporcionar *proxys* significativos sobre la percepción del mercado.

El análisis de sentimientos se ha consolidado como una herramienta fundamental en el ámbito financiero, permitiendo identificar y cuantificar las emociones expresadas en textos, lo cual influye en los precios de las acciones. Este enfoque ha cobrado especial relevancia con el crecimiento de las redes sociales, donde millones de usuarios comparten sus opiniones diariamente, creando una rica fuente de datos para los mercados. Así pues, el análisis de sentimientos se presenta como un enfoque valioso para entender mejor cómo las emociones y percepciones afectan el comportamiento del mercado financiero, ofreciendo herramientas para mejorar la toma de decisiones y anticipar movimientos en un entorno cada vez más influenciado por factores emocionales y sociales.

Ecopetrol, como una de las principales entidades del sector energético en Colombia, se sitúa en el centro del interés financiero, su desempeño está íntimamente relacionado con factores económicos, políticos y sociales que impactan su percepción en el mercado. Estudios previos, como el de Sáenz del Río (2016), han analizado el impacto de los precios del petróleo en el desempeño financiero de Ecopetrol, pero no han considerado el análisis de sentimientos como variable explicativa. Por tanto, resulta relevante desarrollar un indicador de incertidumbre basado en el análisis textual de un gran volumen de noticias y reportes financieros vinculados a Ecopetrol, utilizando las herramientas mencionadas anteriormente.

Estos estudios desafían la teoría clásica de los mercados eficientes, propuesta por Fama (1970), que sostiene que los

agentes actúan racionalmente con información perfecta. Por el contrario, los resultados del análisis de sentimientos indican que las emociones desempeñan un papel crucial en las decisiones de inversión, lo que sugiere que los mercados no siempre operan de manera racional y que las percepciones expresadas en redes sociales y noticias pueden generar volatilidad en los mercados, modificando expectativas y decisiones sobre las empresas.

En este contexto, el trabajo busca llenar este vacío mediante el desarrollo de un indicador de incertidumbre basado en el análisis textual de noticias y reportes financieros, para lo cual se utilizarán los nuevos desarrollos de la inteligencia artificial (IA), las técnicas avanzadas de procesamiento de lenguaje natural (PLN) y los modelos como BERT (*bidirectional encoder representations from transformers*), que perfeccionan el análisis de grandes volúmenes de texto, la identificación de patrones y tendencias lingüísticas, la detección de sentimientos y temas relevantes que pueden influir en la volatilidad del mercado.

Mediante este enfoque que combina el análisis de sentimientos, se busca cuantificar la incertidumbre y su efecto en el comportamiento del mercado. Este esfuerzo proporcionará una herramienta cuantitativa para entender cómo la percepción del mercado, reflejada a través del análisis textual, incide en la volatilidad de los activos de Ecopetrol. Este estudio contribuye a la comprensión más profunda de las dinámicas del mercado y ofrece *insights* valiosos para inversores y analistas financieros.

Revisión de literatura

En la actualidad existen múltiples fuentes digitales de información, cuyo tratamiento adecuado puede utilizarse para generar indicadores indirectos acerca de fenómenos cualitativos como puede ser la incertidumbre (Trust et al., 2023). En este sentido, el desarrollo del índice de incertidumbre de la política económica, denominado EPU, a partir de noticias en más de cinco mil medios de comunicación, fue pionero en identificar esta posibilidad en el ámbito de la incertidumbre (Baker et al., 2016). A partir del mismo, este campo ha sido objeto de análisis y mejoras, sobre todo en lo que respecta al desarrollo de las metodologías de inteligencia artificial, el procesamiento de lenguaje natural y el aprendizaje automático aplicados al análisis de noticias financieras (Baker et al., 2019).

El análisis de sentimientos de mercado está despertando un interés creciente a medida que se desarrollan las metodologías de *machine learning* y *deep learning* (Liapis et al., 2023), y se ha detectado una concentración de análisis en los ámbitos en los que hay una mayor disponibilidad de datos o una mayor base científica, como Estados Unidos, China o la Unión Europea (Heydarian et al., 2025). De esta forma, en este tipo de análisis quedan relegados, en cierto modo, el resto de países, lo que justifica, si cabe, la elección que hemos hecho de trabajar sobre una empresa cotizada y con sede en Colombia.

Los precios del petróleo son especialmente sensibles a las incertidumbres globales (Bloom, 2009) y a los riesgos geopolíticos (Mamman, S. O. et al., 2024), por lo que es aún más necesario analizar los efectos de la incertidumbre sobre las empresas petroleras. Los eventos geopolíticos pueden afectar significativamente la volatilidad del mercado, especialmente en el caso de los futuros del petróleo, que a menudo funcionan como una cobertura contra el riesgo regional para los inversores en acciones (Smales, 2021). De hecho, esta cuestión ya ha sido tratada por trabajos previos para el caso de Colombia (Candelo et al., 2023). Por otra parte, la incertidumbre acerca de las políticas de cambio climático tiene una especial incidencia en las cotizaciones bursátiles de las empresas energéticas en general, y las empresas petroleras en particular (Tedeschi et al., 2024). Para algunos autores (Rao et al., 2024), la mejor forma de intentar visualizar este tipo de incertidumbre, derivada del diseño y la implementación de las políticas de cambio climático, es mediante metodologías de inteligencia artificial, como la que se plantea en este trabajo.

Entre los enfoques más eficientes y generalizados de tratamiento de información masiva de forma automatizada, a través de identificación de noticias como *proxy* de sentimientos, destacan los modelos BERT (Sayeed et al., 2023), como el que nosotros planteamos, y especialmente en el ámbito de análisis de sentimientos financieros. En este sentido, y específicamente aplicado al campo de la incertidumbre, se ha evidenciado que con esta metodología se puede reducir el trabajo de ajuste fino de expertos, que en vez de seleccionar y etiquetar los artículos objeto de tratamiento pueden centrarse en tareas mucho más eficientes como identificar las palabras clave (Trust et al., 2023). También se ha trabajado en la integración y desarrollo de modelos avanzados híbridos específicos de lenguaje masivo de análisis de sentimientos en el ámbito financiero (Shobayo et al., 2024), y específicamente en el ámbito de las predicciones bursátiles (Chen, Boukouvalas et al., 2024; Hao et al., 2025).

Estos nuevos planteamientos permiten capturar con más precisión las sutilezas de los términos del lenguaje financiero, ya que incorporan la capacidad de los métodos tradicionales de tratamiento directo de la información financiera numérica contenida en el volumen de cotización y el precio de las acciones (Liu et al., 2024). Los desarrollos más recientes de análisis de sentimientos en los mercados financieros están incluso utilizando inteligencia artificial generativa (Fatouros et al., 2023), a partir de análisis de noticias en medios de comunicación; pero las limitaciones surgen, ante todo, por el problema de que los nuevos acrónimos o texto

generado cambien el sentido de la fuente original (Delgadillo et al., 2024), lo cual dificulta su aplicación a este dominio y requiere adaptaciones concretas de los programas de tratamiento de lenguaje masivo (Inserte et al., 2024).

La incertidumbre innata de los mercados financieros puede afectar a las cotizaciones de las acciones de diversas formas (Panousi & Papanikolaou, 2012); por ejemplo, provocando una caída de las expectativas de los flujos futuros de caja y un aumento de la aversión al riesgo de los inversores, lo que a su vez eleva la prima de riesgo implícita en la tasa de descuento con la que se valoran los beneficios futuros. De hecho, existe una relación inversa entre incertidumbre y cotizaciones bursátiles (Dzielinski, 2012). Esta relación, además, varía en función del contexto económico (Paule Vianez, 2020).

En este contexto, la dificultad para incorporar adecuadamente el concepto de incertidumbre es elevada. Por ello, en la práctica algunos trabajos suelen incorporar también el análisis de sentimientos de mercado (Chen et al., 2025). Otra posibilidad es seguir la atención que reciben determinadas acciones por parte de los inversores mediante el seguimiento de las búsquedas en internet (Szczygielski et al., 2024). Una dificultad adicional radica en que destacan el papel de los mercados financieros en la amplificación de los efectos adversos de los *shocks* de incertidumbre (Choi & Yoon (2022).

Un punto de inflexión en esta área fue la aparición de la crisis COVID con su consiguiente impacto en las cotizaciones bursátiles, en las que también incidían los análisis de sentimientos planteados por los nuevos desarrollos de la economía conductual y experimental (Biktimirov et al., 2021). Un enfoque complementario es la integración del análisis de sentimientos con metodología de redes neuronales (Das et al., 2024a). Otra área de desarrollo de estas metodologías ha sido el análisis de las criptomonedas (Loginova et al., 2024).

Un reto en el análisis de sentimiento financiero (Du et al., 2024) es incorporar tanto el sentimiento de mercado como el sentimiento del inversor (Yin & Yang, 2023), ya que en los desarrollos más recientes (Du et al., 2024) se ha evidenciado que ambos están fuertemente interrelacionados y que, por tanto, deben incluirse también en los modelos para aumentar su capacidad explicativa. Alternativamente también se ha avanzado en técnicas de análisis de fragmentación de datos para mejorar la precisión de los modelos (Das et al., 2024b).

De cualquier modo, es importante integrar el análisis de sentimientos con los indicadores financieros tradicionalmente utilizados, ya que ello es crítico para mejorar la robustez y capacidad de los modelos en su aplicación real; por ejemplo, la predicción de cotizaciones bursátiles (Siddique et al., 2025). Por último, también se plantea fusionar el modelo BERT con otros modelos complementarios disponibles de inteligencia artificial aplicados al análisis de noticias, con el fin de mejorar sus resultados (Nannepagu et al., 2024).

El enfoque planteado para analizar el impacto de la incertidumbre en las cotizaciones bursátiles de una empresa energética como Ecopetrol es necesario, ya que en este tipo de empresas, la incertidumbre macroeconómica también incide sobre las prácticas de reporte de sus resultados (Rigamonti et al., 2024). Como consecuencia, para este tipo de empresas, la medición del impacto de la incertidumbre

sobre su desempeño a través de su información financiera puede verse distorsionada, en cuanto la propia incertidumbre puede afectar a la calidad de sus resultados reportados (Bermpei et al., 2022).

Así pues, parece que la evaluación del impacto de la incertidumbre sobre una empresa de este tipo es preferible hacerla a través del seguimiento de la cotización bursátil, aun considerando que los efectos aquí señalados también pueden estar incidiendo en ella. A ello, se añade que en los periodos de mayor incertidumbre, los mercados financieros tienden a penalizar bursátilmente a las empresas de carácter cíclico, como las empresas petroleras (Atif M. et al., 2022), entre otras cosas, porque los propios precios del petróleo pueden estar descontando ya esta situación adversa.

En este contexto, algunas empresas petroleras tienden a reducir discrecionalmente sus beneficios (Chauhan & Jaiswall, 2023), ralentizando sus ritmos de extracción de sus reservas o gestionando los gastos discrecionales, para preservar estas ganancias para los periodos en los que la incertidumbre es menor y, por tanto, el impacto positivo de sus beneficios sobre la cotización de sus acciones es potencialmente mayor (Kjærland et al., 2021). En efecto, las políticas de inversión de las empresas petroleras son una de las vías que mejor reflejan la incertidumbre (Kellogg, 2014). De hecho, existe una correlación inversa entre la incertidumbre y las inversiones en empresas petroleras a nivel internacional (Chen, Dong et al., 2024), lo que resulta paradójico, dado que en este tipo de estas empresas las necesidades de inversión suelen ser continuas en tiempo para poder desarrollar su actividad.

Metodología

Los modelos de lenguaje natural

El análisis de sentimientos es una técnica de procesamiento de lenguaje natural (PLN), que permite identificar y cuantificar las emociones y opiniones expresadas en textos. Este enfoque se ha vuelto esencial en varios campos, especialmente en el análisis de datos financieros, redes sociales y medios de comunicación. El objetivo principal del análisis de sentimientos es clasificar un texto en categorías como positivo, negativo o neutral, y evaluar la intensidad del sentimiento expresado. Tradicionalmente, este análisis se realizaba utilizando métodos lexicográficos, pero con el avance de las arquitecturas de aprendizaje profundo, como BERT, se han logrado mejoras significativas en la precisión y profundidad de los análisis.

BERT (*bidirectional encoder representations from transformers*), introducido por Jacob Devlin y su equipo de investigación en Google AI, en el artículo titulado “BERT: Pre-training of deep bidirectional transformers for language understanding”, publicado en el 2018, ha revolucionado el análisis de sentimientos al capturar de manera más precisa las relaciones contextuales dentro de los textos. Gracias a su arquitectura bidireccional, BERT procesa simultáneamente la información desde ambos lados de una secuencia de palabras, lo que le permite comprender mejor el contexto global de una oración o fragmento de texto. El modelo se

entrena utilizando técnicas como el modelado de lenguaje enmascarado (MLM, por sus siglas en inglés), donde se ocultan algunas palabras del texto y el modelo debe predecirlas, y la predicción de la siguiente oración (NSP, por sus siglas en inglés), que evalúa la relación entre oraciones consecutivas.

Estas capacidades hacen de BERT una herramienta poderosa para el análisis de sentimientos, ya que no solo identifica el sentimiento predominante en un texto, sino que también puede ajustar su comprensión del texto según los matices contextuales. A través del ajuste fino con datos etiquetados de sentimientos, BERT genera distribuciones de probabilidad sobre las posibles categorías de sentimiento, evaluando la confianza de sus predicciones. Esto permite que las aplicaciones de análisis de sentimientos sean más precisas y puedan capturar la complejidad emocional en diferentes dominios textuales, desde opiniones de usuarios hasta reportes financieros. Este esquema se resume en la figura 1.

XLM-RoBERTa es un modelo de lenguaje multilingüe que ha sido entrenado en un extenso corpus de datos en varios idiomas. Su diseño permite que el modelo aprenda patrones comunes entre diferentes lenguas, lo que mejora su rendimiento en tareas de procesamiento de lenguaje natural (PLN) en contextos multilingües. Esto lo convierte en una herramienta valiosa para aplicaciones que requieren análisis en múltiples idiomas, como traducción automática y clasificación de textos. XLM-RoBERTa fue seleccionado por su capacidad para manejar textos multilingües, lo que es crucial dado el contexto global de Ecopetrol (Conneau et al., 2020).

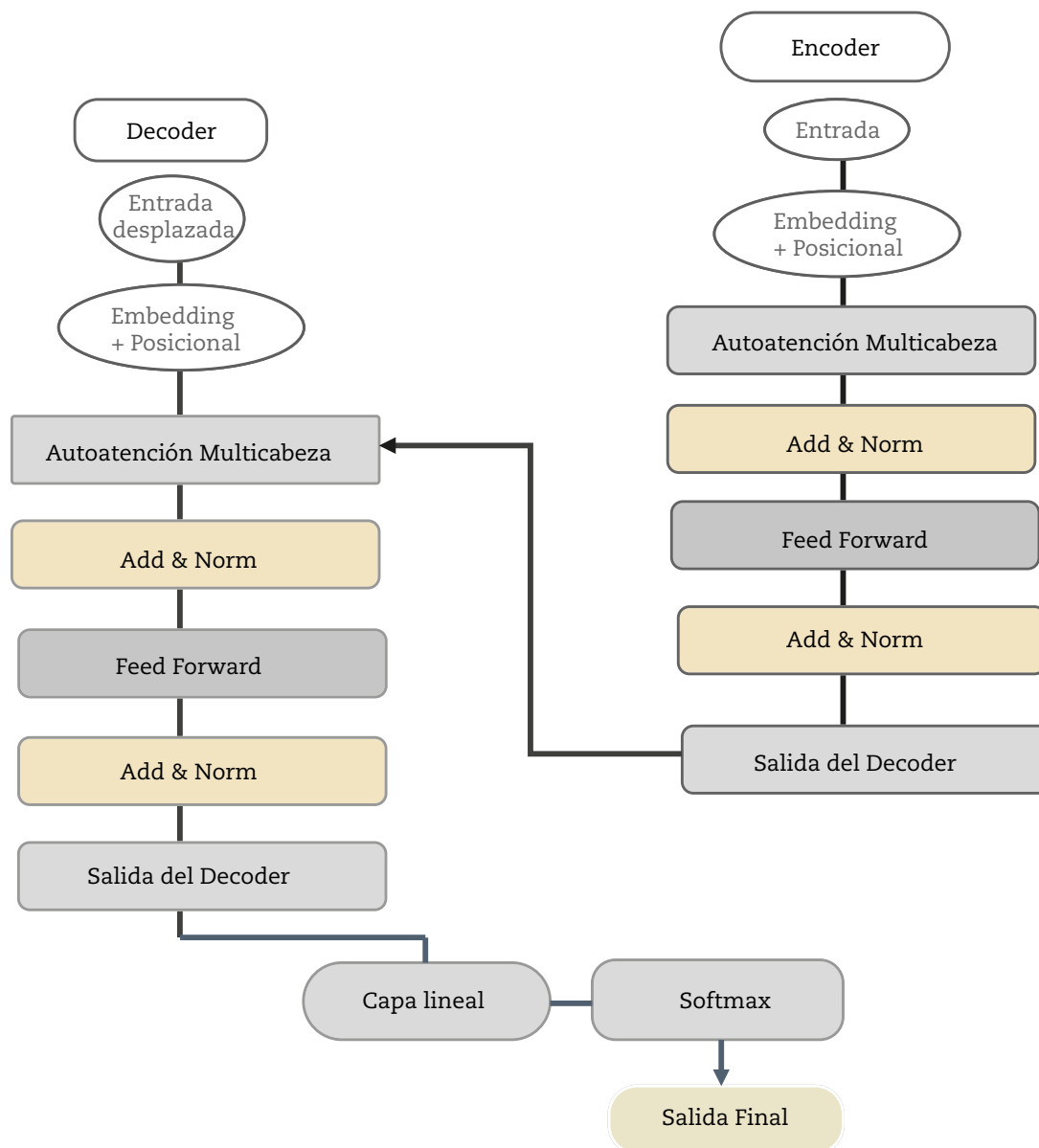
BETO, por otra parte, es una variante de BERT optimizada para el idioma español. Este modelo ha sido pre-entrenado con un gran volumen de texto en español, lo que le permite captar las particularidades lingüísticas y culturales del idioma. Su enfoque especializado lo hace ideal para tareas de análisis de sentimientos y comprensión del lenguaje en contextos hispanohablantes.

DistilBERT es una versión más ligera y eficiente de BERT, diseñada para mantener la mayor parte del rendimiento del modelo original mientras reduce el tamaño y los requisitos computacionales. Esta reducción hace que DistilBERT sea especialmente útil en aplicaciones donde la velocidad y la eficiencia son críticas, como en dispositivos móviles o entornos con recursos limitados.

FinBERT es un modelo específico para el sector financiero, pre-entrenado con textos relacionados con comunicaciones financieras. Su especialización le permite realizar análisis de sentimientos con alta precisión en contextos financieros, superando a modelos genéricos como BERT. Esto lo convierte en una herramienta esencial para analistas e investigadores que necesitan evaluar el sentimiento del mercado a partir de noticias y reportes financieros. FinBERT, en cambio, fue elegido por su especialización en el análisis de sentimientos en el dominio financiero (Araci, 2019).

GPT-2, por su parte, es un modelo generativo que se destaca en la creación de texto coherente a partir de una entrada inicial. Aunque no está diseñado específicamente para el análisis financiero, su capacidad para generar contenido puede ser aprovechada para crear informes o resúmenes financieros automatizados, facilitando la comunicación en este ámbito.

Figura 1. Arquitectura modelos de transformer



Fuente: elaboración propia, basada en Devlin et al. (2018).

DeBERTa introduce innovaciones arquitectónicas que mejoran la comprensión del contexto y la semántica del texto. Este modelo ha demostrado ser eficaz en diversas tareas del PLN al integrar mecanismos avanzados de atención, lo que le permite captar mejor las relaciones entre palabras y frases dentro del texto. Su flexibilidad lo hace adecuado para aplicaciones tanto generales como específicas, incluyendo el análisis financiero.

Datos

Los datos utilizados en este estudio consisten principalmente en textos provenientes de noticias y reportes financieros relacionados con Ecopetrol, tanto en español como en

inglés. Estos documentos se recopilieron y se sometieron a un proceso de limpieza y preprocesamiento para eliminar el ruido textual (como menciones, enlaces y caracteres especiales) y preparar el contenido para su análisis con modelos de lenguaje avanzados como FinBERT, BETO, XLM-RoBERTa, DeBERTa y otros.

A partir de estos textos, se llevó a cabo un etiquetado manual de 979 titulares y 1421 documentos financieros, y se clasificaron según variables como polaridad, subjetividad, confianza, novedad y certeza, y luego se integraron en un índice compuesto denominado "sentimiento final". Además, se aplicó una medición de la entropía sobre las distribuciones mensuales de polaridad para cuantificar el grado

de incertidumbre en el sentimiento expresado en los textos: una entropía alta indica mayor dispersión y, por tanto, mayor incertidumbre. Por último, estos datos cualitativos se complementaron con las variables financieras como el precio ajustado de cierre de Ecopetrol, la volatilidad estimada mediante modelos GARCH, la volatilidad implícita, la tasa representativa del mercado (TRM), el índice de precios al productor (IPP) y el índice COLCAP, con el fin de analizar cómo las variaciones en el sentimiento y su incertidumbre se relacionan con el comportamiento de los activos de la compañía.

Tipo, alcance y diseño

La investigación se aborda desde la metodología mixta, al tomar las variables cualitativas para procesarlas mediante un enfoque cuantitativo novedoso, basado en el enfoque correlacional propuesto por Xhao et al. (2021), quienes utilizaron modelos de lenguaje para el análisis de sentimientos en el dominio financiero.

Procedimiento

Se construyó un modelo de análisis con las siguientes funciones para optimizar la limpieza y el preprocesamiento:

- **Función *clean tweet* (*tweet*):** elimina menciones, enlaces y caracteres especiales que hay en los *tweets*, además de convertir todo el texto a minúsculas. Este proceso reduce el ruido en los datos y asegura que los modelos trabajen con un texto más uniforme y relevante.
- **Función *preprocess text* (*text*):** realiza una limpieza adicional del texto eliminando enlaces y puntuación, además de tokenizar el texto en palabras individuales. Este preprocesamiento es utilizado tanto para el análisis de sentimientos basado en léxico como para el análisis mediante modelos BERT.

Los textos en diferentes idiomas se preprocesaron con procesadores de texto multilingües, que permiten el análisis directo de datos en múltiples idiomas sin necesidad de traducción a un idioma común. Este enfoque aprovecha herramientas avanzadas de procesamiento de lenguaje natural (PLN), como la API de Google Natural Language, que soporta el análisis de sentimientos, entidades y categorización de texto en varios idiomas.

El proceso de preprocesamiento se llevó a cabo siguiendo estos pasos:

1. **Carga de modelos:** se cargaron los modelos pre-entrenados, junto con sus respectivos tokenizadores. Estos modelos están diseñados para realizar tareas de clasificación de sentimientos sobre textos.
2. **Carga de datos:** se importó el conjunto de datos desde un archivo Excel, que contiene los textos que se van a analizar. Se crearon columnas adicionales en el DataFrame para almacenar los resultados de las predicciones de sentimiento generadas por cada modelo.
3. **Limpieza y tokenización:** se aplicaron las funciones de limpieza y preprocesamiento. Esto incluyó la eliminación de menciones, enlaces y caracteres especiales, así como la conversión de textos a

minúsculas. Luego se tokenizó cada texto en palabras individuales para prepararlos para el análisis.

Análisis de datos

Análisis de sentimiento e impacto con modelos de BERT ajustados

En este proyecto, se emplean modelos como FinBERT, XLM-RoBERTa, BETO y DeBERTa para realizar análisis de sentimientos sobre noticias, reportes y documentos de Ecopetrol. Los modelos proporcionan una polaridad continua en un rango de $[-1, 1]$, donde -1 indica un sentimiento extremadamente negativo y 1 un sentimiento extremadamente positivo. Esta polaridad continua nos permite captar con precisión las fluctuaciones emocionales en el texto y su impacto en los precios de las acciones de la compañía.

El objetivo de usar modelos en varios idiomas es capturar los matices y la percepción internacional hacia Ecopetrol. El uso de modelos como XLM-RoBERTa y DeBERTa, entrenados en múltiples lenguas, permite identificar cómo perciben la compañía los diferentes mercados globales. BETO, por otra parte, se centra en el análisis detallado de textos en español.

Se utilizaron en el análisis los siguientes modelos:

- **'yiyanghkust/finbert-tone':** especializado en análisis de tono financiero, entrenado para capturar los matices emocionales específicos de reportes financieros y noticias relacionadas con mercados financieros.
- **'xlm-roberta-base':** un modelo multilingüe que permite capturar percepciones internacionales en diferentes lenguajes. Su capacidad multilingüe es clave para comprender cómo las noticias internacionales afectan la imagen y percepción de Ecopetrol.
- **'dccuchile/bert-base-spanish-wwm-cased (BETO)':** un modelo entrenado específicamente para el español, utilizado para capturar los matices del lenguaje local en reportes y noticias en español, lo que permite un análisis profundo de textos regionales.
- **'microsoft/deberta-base':** un modelo robusto que mejora la comprensión contextual, y permite capturar detalles más complejos del lenguaje financiero en múltiples idiomas.
- **'nlp-town/bert-base-multilingual-uncased-sentiment':** usado para captar un análisis de sentimientos global en un contexto multilingüe.
- **'gpt2':** utilizado como modelo complementario para el análisis de estructuras narrativas y la captura de matices emocionales más complejos en los textos analizados.

Fine-tuning supervisado para análisis de sentimientos

El análisis de sentimientos supervisado se implementó mediante el ajuste fino (*fine-tuning*) de varios modelos de lenguaje previamente entrenados, utilizando una muestra etiquetada manualmente. Esta muestra se seleccionó de manera aleatoria de documentos financieros y reportes de

Ecopetrol, con el objetivo de capturar matices importantes sobre el sentimiento hacia la compañía.

Se calcula el tamaño de la muestra con un nivel de confianza del 99% y un margen de error del 5%; el tamaño de la muestra necesario es $n \approx 664$; utilizamos 979 muestras etiquetadas de titulares de noticias para maximizar estos parámetros.

El proceso de etiquetado manual incluyó la clasificación de los textos en categorías que reflejan la polaridad del contenido, la subjetividad, la confianza, la novedad y la certeza. Esto permitió entrenar los modelos de manera precisa, considerando las siguientes ponderaciones para el *sentimiento final* (SF):

$$SF = 0.4Pol + 0.2Subj + 0.2Conf + 0.1Nov + 0.1Cert \quad (1)$$

Donde:

- Pol: representa la polaridad del texto, es decir, si el contenido es positivo, negativo o neutral.
- Subj: mide la subjetividad, e indica el grado de opinión personal frente a la información objetiva.
- Conf: evalúa la confianza, y determina la seguridad con la que se presentan las afirmaciones.

- Nov: indica la novedad, y muestra cuán innovador o reciente es el contenido.
- Cert: captura la certeza, y evalúa el grado de claridad en la información presentada.

La figura 2 resume las distribuciones de polaridad encontradas mediante la aplicación de la metodología descrita anteriormente.

Análisis de sentimientos e impacto supervisado

Paralelamente, se llevó a cabo un análisis supervisado de sentimiento e impacto sobre una gran cantidad de documentos financieros. En este caso, se aplicó un etiquetado manual, para realizar predicciones continuas sobre el sentimiento y el impacto potencial en los mercados, con 1421 muestras etiquetadas. La figura 3 describe estos resultados.

El análisis se efectuó en lotes grandes de documentos, mediante el procesamiento en paralelo con infraestructura de GPU para mejorar la eficiencia. Los resultados del análisis de sentimientos e impacto se guardaron y se utilizaron para modelar la percepción del mercado sobre la compañía, integrándose con otros enfoques de análisis de volatilidad y riesgo financiero.

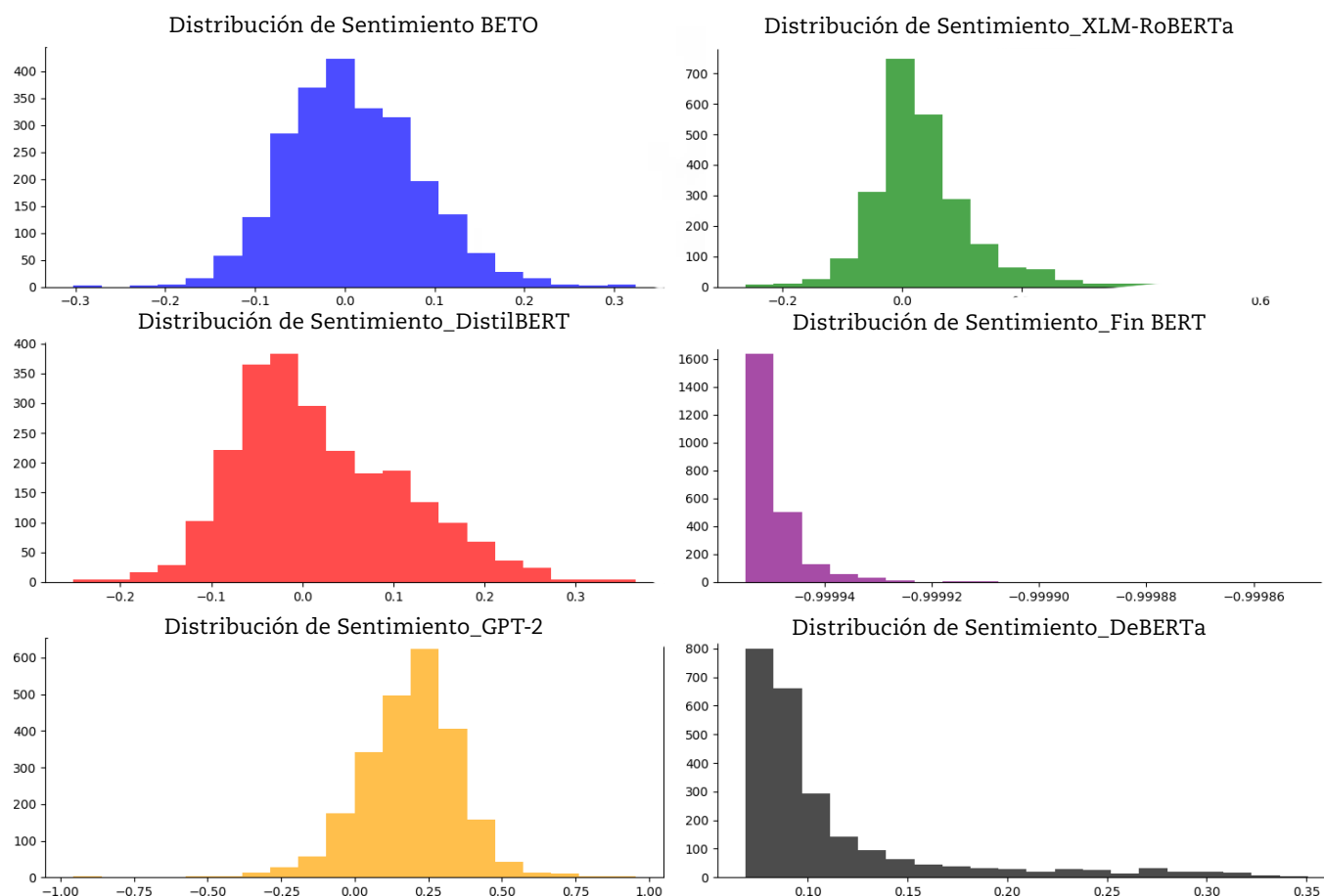


Figura 2. Distribuciones de las polaridades de los modelos basados en BERT

Fuente: elaboración propia.

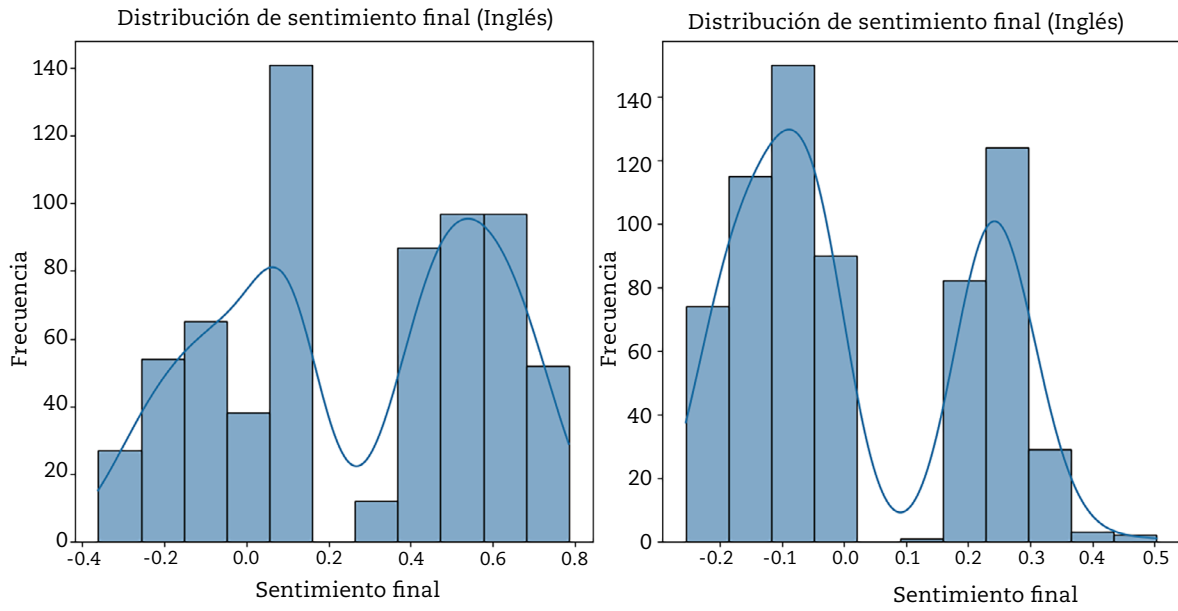


Figura 3. Distribución de sentimiento final

Fuente: elaboración propia.

Optimización bayesiana y mezcla gaussiana para ensemble

En este trabajo se combinan las predicciones de dos grupos de modelos de análisis de sentimientos con el fin de encapsular los matices de cada modelo. El primer grupo incluye Fin-BERT, GPT-2 y DeBERTa. El segundo grupo incluye XLM, BETO y DistilBERT.

Para cada grupo, se utiliza una mezcla gaussiana (GMM) para modelar la distribución subyacente de las predicciones y posteriormente se optimizan los pesos con un enfoque de optimización bayesiana. A continuación, se presenta el desarrollo matemático.

a. Predicciones de los modelos

Denotemos las predicciones de cada modelo de sentimiento en el primer grupo como

$$\hat{y}_{FinBERT} = X_1, \hat{y}_{GPT-2} = X_2, \hat{y}_{DeBERTa} = X_3 \quad (2)$$

El vector de predicciones concatenado para el primer grupo es:

$$X_{Grupo1} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

Y para el segundo grupo:

$$\hat{y}_{XLM} = X_4, \hat{y}_{BETO} = X_5, \hat{y}_{DistilBERT} = X_6 \quad (3)$$

El vector de predicciones concatenado para el segundo grupo es:

$$X_{Grupo2} = \begin{bmatrix} X_4 \\ X_5 \\ X_6 \end{bmatrix}$$

b. Normalización y mezcla gaussiana

Las predicciones de ambos grupos se normalizan utilizando una transformación estándar para asegurar que todas tengan la misma escala:

$$X_{scaled} = \frac{X - \mu_X}{\sigma_X} \quad (4)$$

Donde μ_X y σ_X son la media y desviación estándar de las predicciones concatenadas de cada grupo.

Luego, se aplica una mezcla gaussiana (GMM) para estimar la distribución subyacente de las predicciones. La función de densidad de probabilidad para un modelo GMM con K componentes es:

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x | \mu_k, \Sigma_k) \quad (5)$$

Donde: π_k es el peso de la k -ésima componente.

$\mathcal{N}(x | \mu_k, \Sigma_k)$ es la distribución gaussiana con media μ_k y covarianza Σ_k .

El número de componentes K se selecciona optimizando el criterio de información bayesiano (BIC, por sus siglas en inglés). Las ponderaciones asignadas por el GMM a cada predicción se denotan como w_{GMM} , y la predicción ponderada final para cada grupo es:

Para el grupo 1:

$$\hat{y}_{GMM, grupo1} = \sum_{k=1}^K w_{GMM, grupo1, k} X_k \quad (6)$$

Para el grupo 2:

$$\hat{y}_{GMM, grupo2} = \sum_{k=1}^K w_{GMM, grupo2, k} X_{k+3} \quad (7)$$

c. Optimización bayesiana

Para encontrar los pesos óptimos entre los modelos del grupo 1, se utiliza la optimización bayesiana. Denotemos los pesos de cada modelo del grupo 1 como w_1, w_2, w_3 , donde las restricciones son:

$$w_1 + w_2 + w_3 = 1 \quad (8)$$

El objetivo es minimizar el error absoluto medio (MAE, por sus siglas en inglés) entre la predicción combinada y los valores reales. La predicción ponderada total para el grupo 1 se define como:

$$\hat{y}_{final, grupo1} = \sum_{i=1}^3 w_i \hat{y}_{modelo_i} + \hat{y}_{GMM, grupo1} \quad (9)$$

Para penalizar si algún peso excede el 60%, se introduce un término de penalización P , que es la suma de los cuadrados del exceso:

$$P = (\max(0, w_1 - 0.6))^2 + (\max(0, w_2 - 0.6))^2 + (\max(0, w_3 - 0.6))^2 \quad (10)$$

El objetivo final es minimizar la siguiente función de costo para el grupo 1:

$$MAE_{grupo1} + P = \frac{1}{N} \sum_{j=1}^N |\hat{y}_{final, grupo1, j} - y_{true, j}| + P \quad (11)$$

De manera similar, para el grupo 2 denotamos los pesos de cada modelo como w_4, w_5, w_6 , donde las restricciones son:

$$w_4 + w_5 + w_6 = 1 \quad (12)$$

El objetivo es minimizar el error absoluto medio (MAE) entre la predicción combinada y los valores reales. La predicción ponderada total para el grupo 2 se define como:

$$\hat{y}_{final, grupo2} = \sum_{j=1}^3 w_j + 3\hat{y}_{modelo_{j+3}} + \hat{y}_{GMM, grupo2} \quad (13)$$

Al igual que en el grupo 1, se introduce un término de penalización P :

$$P = (\max(0, w_4 - 0.6))^2 + (\max(0, w_5 - 0.6))^2 + (\max(0, w_6 - 0.6))^2 \quad (14)$$

El objetivo final es minimizar la siguiente función de costo para el grupo 2:

$$MAE_{grupo2} + P = \frac{1}{N} \sum_{j=1}^N |\hat{y}_{final, grupo2, j} - y_{true, j}| + P \quad (15)$$

d. Combinación final de ponderaciones

Tras aplicar la optimización bayesiana, los pesos óptimos encontrados para el grupo 1 son:

$$w^*_{FinBERT}, w^*_{GPT-2}, w^*_{DeBERTa} \quad (16)$$

La combinación final de la predicción para el grupo 1 es entonces:

$$\hat{y}_{final, grupo1} = w^*_{FinBERT} \hat{y}_{FinBERT} + w^*_{GPT-2} \hat{y}_{GPT-2} + w^*_{DeBERTa} \hat{y}_{DeBERTa} + \hat{y}_{GMM, grupo1} \quad (17)$$

Tras aplicar la optimización bayesiana, los pesos óptimos encontrados para el grupo 2 son:

$$w^*_{XLM}, w^*_{BETO}, w^*_{DistilBERT} \quad (18)$$

La combinación final de la predicción para el grupo 2 es entonces:

$$\hat{y}_{final, grupo2} = w^*_{XLM} \hat{y}_{XLM} + w^*_{BETO} \hat{y}_{BETO} + w^*_{DistilBERT} \hat{y}_{DistilBERT} + \hat{y}_{GMM, grupo2} \quad (19)$$

La figura 4 presenta los resultados de la distribución de la predicción de sentimiento utilizando titulares en español.

Por su parte, la figura 5 presenta los resultados de la distribución de la predicción de sentimiento utilizando titulares en inglés.

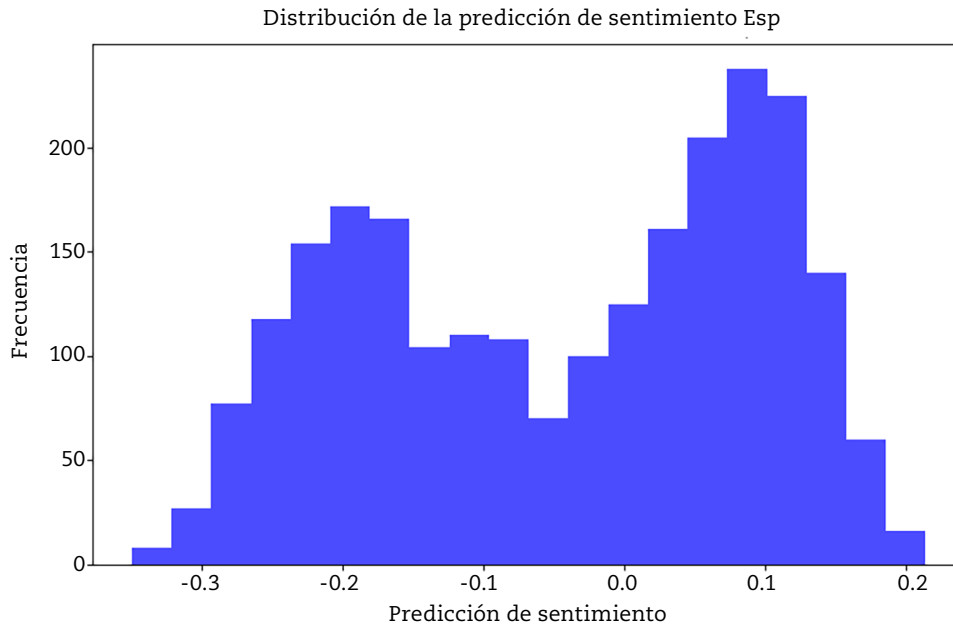


Figura 4. Distribución de la predicción de sentimiento titulares en español
Fuente: elaboración propia.

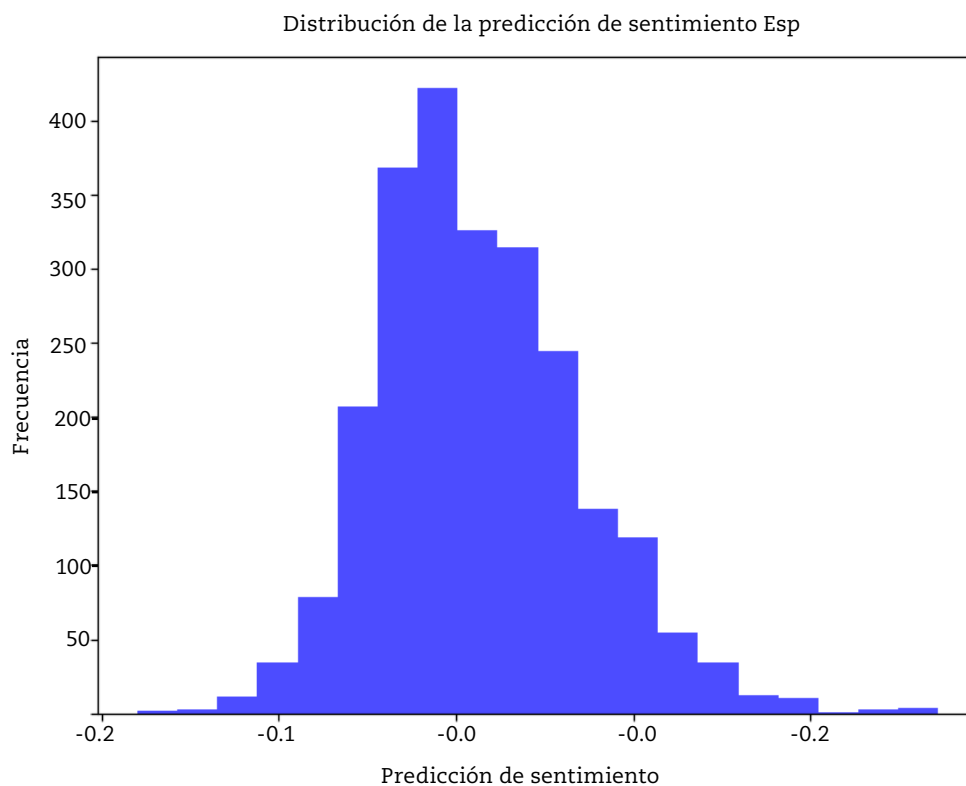


Figura 5. Distribución de la predicción de sentimiento titulares en inglés

Fuente: elaboración propia.

Cálculo de la entropía

La entropía continua se utiliza para cuantificar la incertidumbre en una distribución de polaridades mensuales. La entropía $H(X)$ de una variable aleatoria X con función de densidad de probabilidad $p(x)$ se define como:

$$H(X) = - \int_{-\infty}^{\infty} p(x) \log(p(x)) dx \quad (20)$$

Esta fórmula permite calcular la entropía de las polaridades mensuales, donde $p(x)$ representa la función de densidad de probabilidad asociada a las polaridades.

La entropía proporciona una medida de la incertidumbre asociada a la distribución de polaridades. Un valor alto de entropía indica una mayor diversidad en las polaridades, mientras que un valor bajo sugiere una concentración más fuerte de las opiniones.

Resultados

El preprocesamiento también implicó la selección de las variables relevantes para el análisis, como sentimiento, impacto, magnitud, persistencia, velocidad, propagación y severidad, resultado del modelo.

Las métricas proporcionadas destacan el rendimiento de varios modelos de lenguaje en diferentes dominios. BETO, con un F1 score de 81.1%, es eficaz en tareas generales en

español. RoBERTa, con un F1 score de 88.5%, sobresale en inglés para tareas generales, mientras que XLM-RoBERTa ofrece flexibilidad multilingüe con un F1 de 79.2% en español. En el ámbito financiero, FinBERT destaca con una alta precisión (F1 de 91.4%) en el análisis de textos financieros en inglés, seguido de FinancialBERT con un rendimiento promedio de 90%. Por su parte, Stock Market RoBERTa muestra un F1 entre 90% y 92%, optimizado específicamente para el análisis de mercado. Estos modelos varían en especialización, pero los diseñados para finanzas muestran mejor rendimiento en ese dominio.

En la figura 6 observamos que la entropía ING*(noticias en inglés) y ESP(noticias en español) muestran patrones de fluctuación distintos pero relacionados. La entropía ING tiende a tener valores más altos y una mayor volatilidad que la entropía ESP. Estas diferencias enriquecen el análisis al proporcionar dos perspectivas de la incertidumbre en el mercado de Ecopetrol. Los picos y valles en ambas medidas de entropía a menudo coinciden con cambios significativos en la varianza y el precio de Ecopetrol, lo que sugiere una fuerte interrelación entre estas métricas.

La figura 7 complementa este análisis al mostrar medidas adicionales de entropía (Esp y Engl) junto con otras métricas de volatilidad e incertidumbre. Aunque la escala temporal no es directamente comparable con la figura 6, podemos observar que los picos en la entropía Esp y Engl a menudo coinciden con aumentos en la media volatilidad GARCH y la

media volatilidad implícita. Esta convergencia de diferentes métricas enriquece nuestra comprensión de los periodos de alta incertidumbre en el mercado de Ecopetrol.

La figura 8 muestra la relación entre el precio de una acción normalizado y los sentimientos provenientes de distintas fuentes a lo largo del tiempo.

Se observan acumulaciones de puntos clave en diferentes periodos que coinciden con eventos relevantes. A inicios del 2020, la caída abrupta del precio de la acción correspon-

de probablemente al inicio de la pandemia de COVID-19, cuando los mercados globales sufrieron fuertes caídas debido a la incertidumbre económica, reflejada en una alta densidad de puntos rojos asociados a titulares negativos. A mediados del 2020, el inicio de la recuperación coincide con estímulos económicos, avances en el desarrollo de vacunas y la adaptación de las empresas, lo que se refleja en puntos naranjas y azules relacionados con calificaciones de riesgo y reputación.

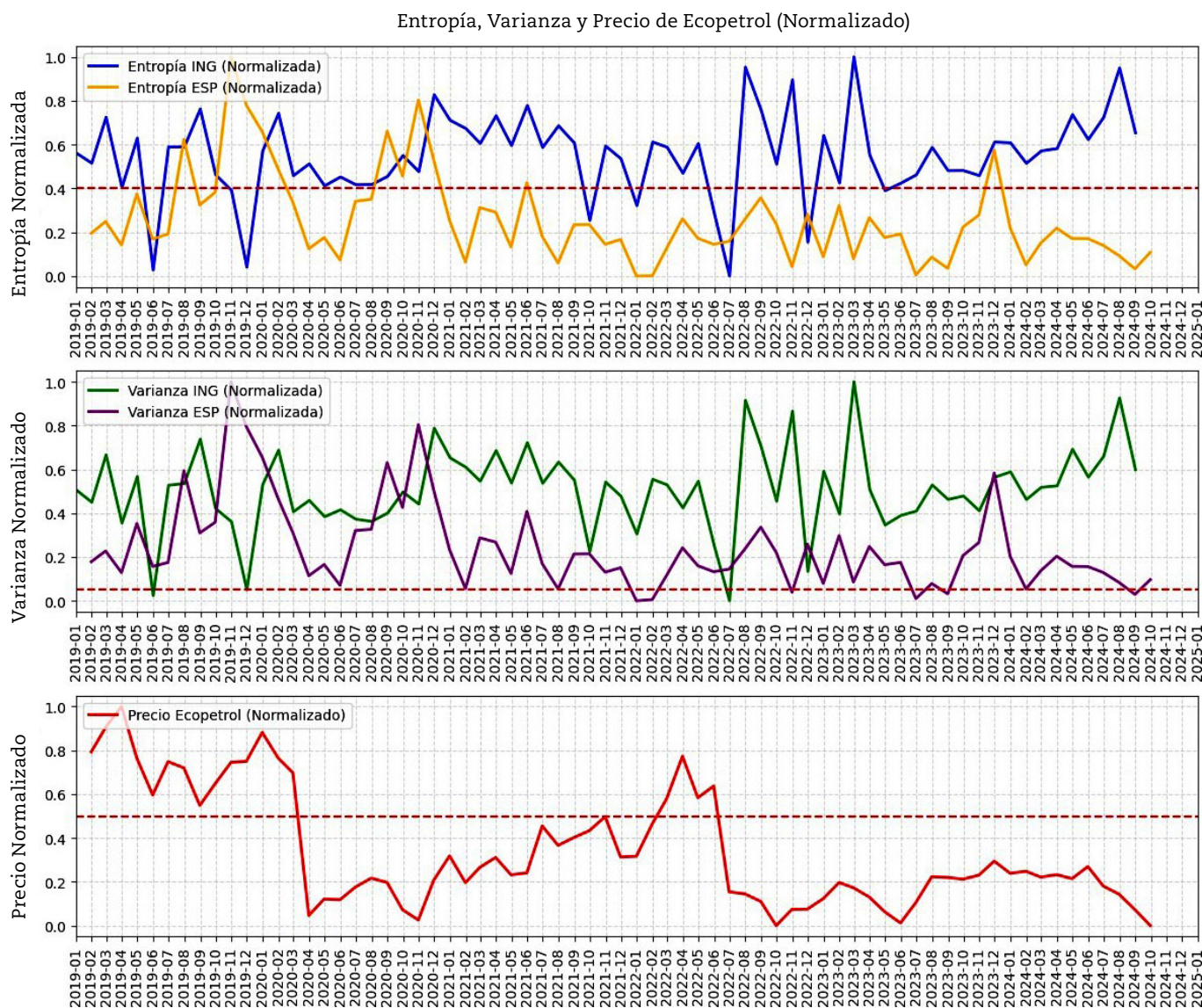


Figura 6. Entropía titulares vs. Ecopetrol

Fuente: elaboración propia.

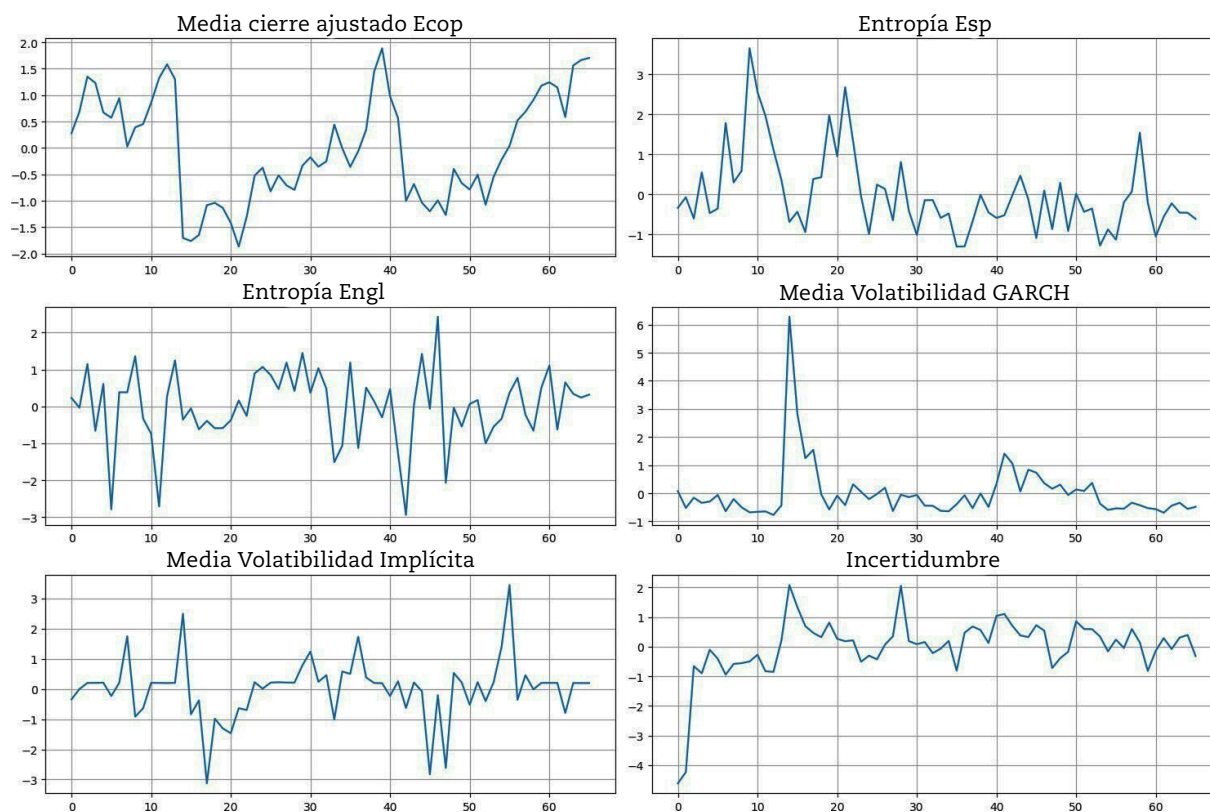


Figura 7. Cierre ajustado Ecopetrol, entropía titulares español, entropía titulares inglés, media volatilidad GARCH, volatilidad implícita, incertidumbre

Fuente: elaboración propia.

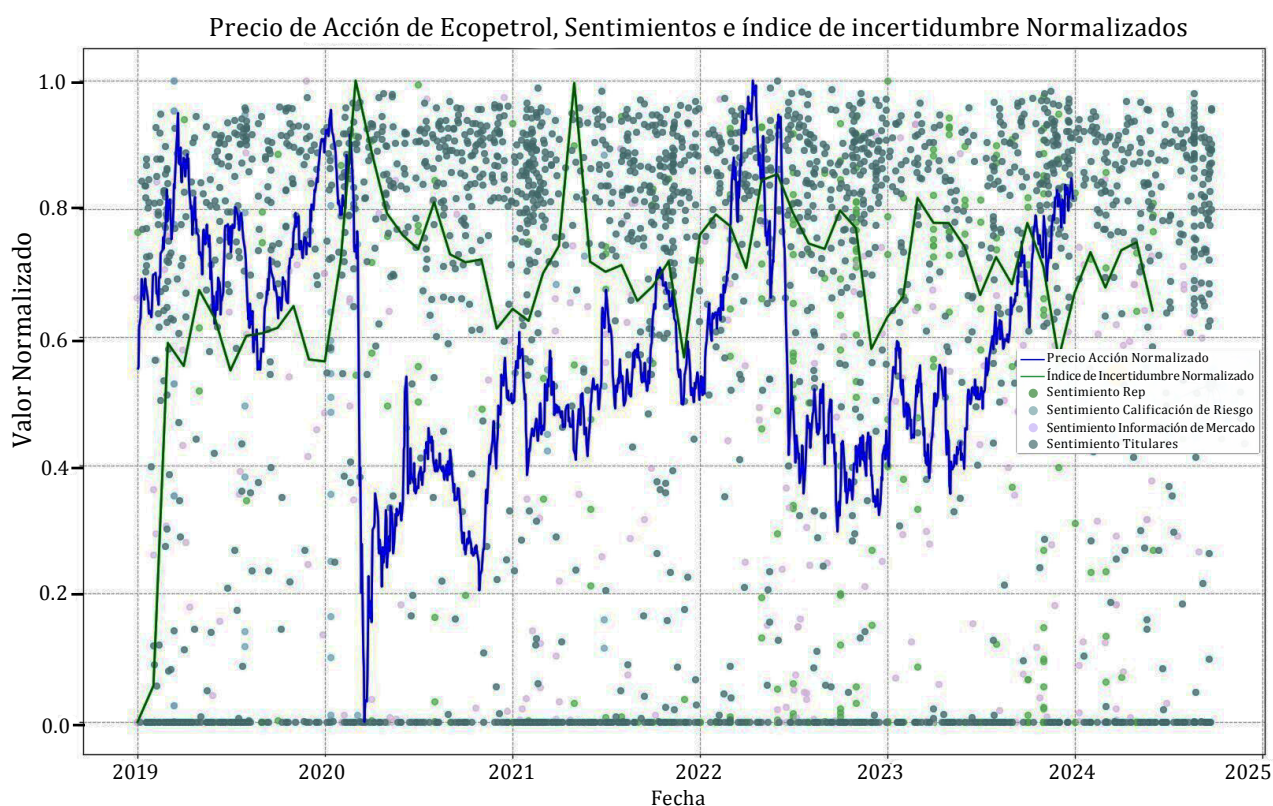


Figura 8. Cierre ajustado Ecopetrol, entropía titulares español, entropía titulares inglés, media volatilidad del COLCAP, índice de precios al productor, tasa representativa del mercado

Fuente: elaboración propia.

En el 2022, los picos en el precio podrían vincularse con eventos como la recuperación pospandemia, resultados financieros positivos o ajustes en las tasas de interés, acompañados de titulares frecuentes y evaluaciones optimistas. Por último, entre finales del 2023 y principios del 2024, el crecimiento sostenido del precio parece reflejar optimismo sobre el futuro económico o innovaciones sectoriales, destacándose nuevamente una alta densidad de puntos rojos relacionados con titulares.

Este análisis sugiere que los sentimientos, especialmente los relacionados con titulares y calificaciones, tienen una correlación significativa con los movimientos del precio de la acción.

Al proyectar los datos sobre las dos primeras componentes principales, como se observa en la figura 9, se aprecia una estructura curva que refleja una relación no lineal entre las variables. Los clústeres identificados presentan una progresión a lo largo de la componente 1, lo que destaca su relevancia en la diferenciación de los grupos.

Además, se identifica una mayor concentración de datos en ciertos extremos de la curva, lo que podría señalar grupos más recurrentes o relevantes en esas áreas. Este análisis confirma que las componentes principales capturan de manera efectiva la variación y estructura subyacente de los datos, facilitando su interpretación y reducción dimensional.

El análisis de las cargas de las variables en las componentes principales de la figura 10, muestra que la componente 1 está influenciada principalmente por las variables como sentimiento, con una carga positiva significativa; mientras que las variables como velocidad, propagación e impacto presentan cargas negativas, indicando una relación inversa. Por otra parte, la componente 2 está dominada por la variable persistencia, con una carga positiva muy destacada, mientras que las demás variables tienen un impacto menor en esta dimensión. Este patrón sugiere que la componente 1 captura una combinación de factores relacionados con la intensidad del fenómeno, mientras que la componente 2 se centra principalmente en su duración o persistencia.

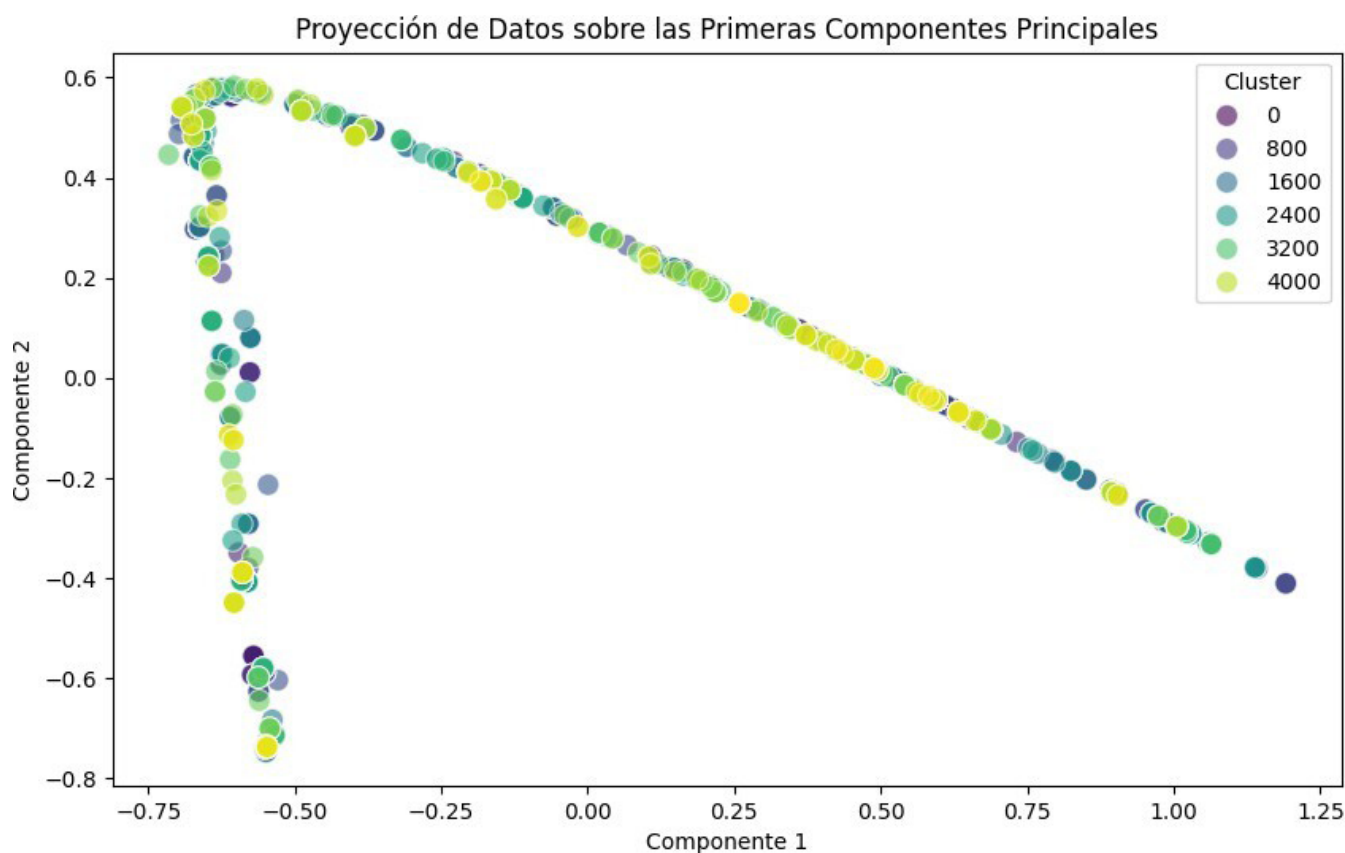


Figura 9. Primera proyección de componentes
Fuente: elaboración propia.

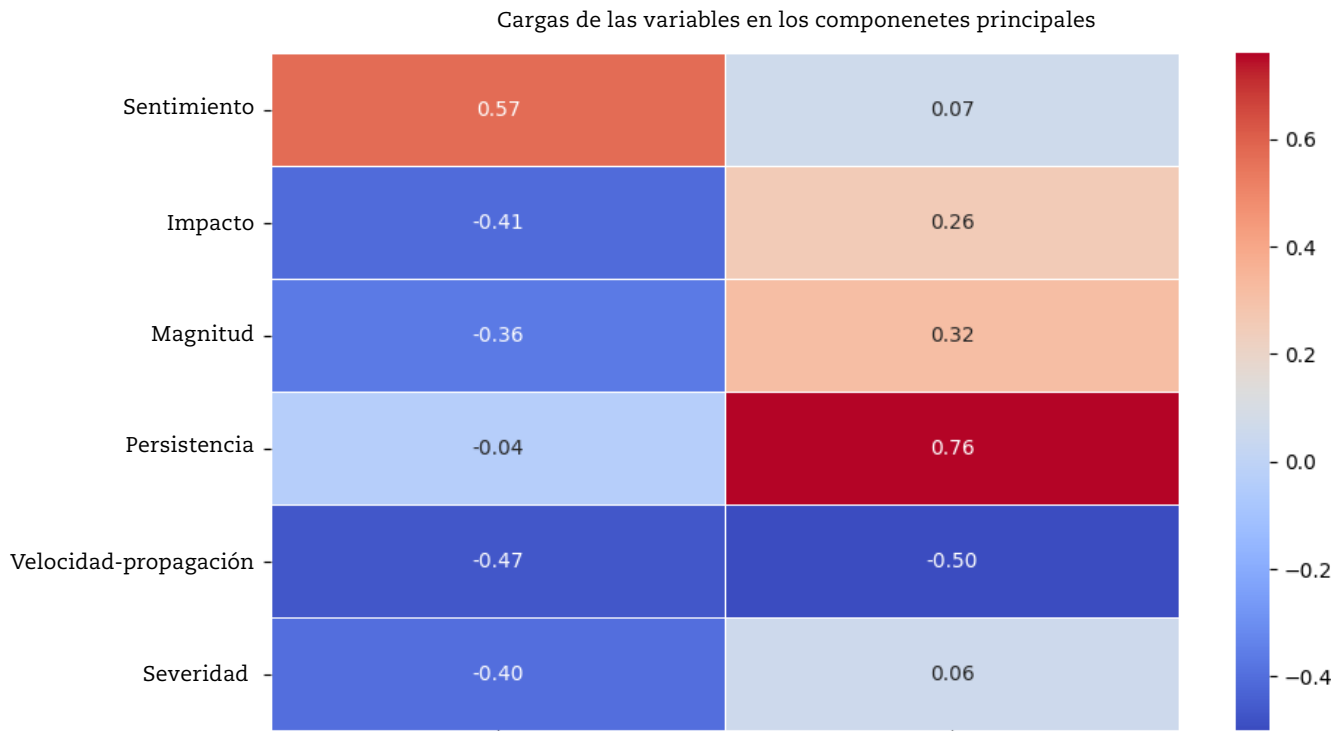


Figura 10. Carga de las variables
Fuente: elaboración propia.

La combinación de estas dos tablas proporciona una visión más completa de la dinámica del mercado. Mientras que la figura 6 ofrece una perspectiva a largo plazo y normalizada, la figura 7 muestra medidas más específicas que pueden capturar eventos o cambios a corto plazo. Esta riqueza de información permite un análisis más robusto de los patrones de entropía y su relación con otras métricas de mercado, lo que facilita una mejor comprensión de los periodos de estabilidad e inestabilidad en el comportamiento de Ecopetrol.

La evolución del precio normalizado muestra variaciones significativas desde el 2019 hasta el 2024, como se ve en la figura 11, con un descenso marcado en ciertos periodos (por ejemplo, a mediados del 2020), seguido de una recuperación gradual. Este comportamiento puede reflejar eventos económicos globales o específicos de la compañía, como cambios en el mercado energético.

Este indicador presenta dispersión constante, con algunos valores más elevados hacia los últimos años (2023-2024). Esto podría sugerir una mejora en la percepción de los representantes o noticias relacionadas. Las observaciones son más escasas y se concentran en niveles específicos de sentimiento, lo que indica menor frecuencia o relevancia de este indicador en comparación con otros. Muestra un patrón más uniforme a lo largo del tiempo, lo que indica una influencia

constante de la información del mercado en el análisis de la acción. Este indicador tiene una alta densidad de datos a lo largo del tiempo, lo que muestra una variabilidad significativa y, probablemente, un fuerte impacto de las noticias en el mercado.

En el análisis de la figura 11, el precio ajustado de cierre de Ecopetrol revela una caída significativa a principios del 2020, probablemente como consecuencia de la pandemia de COVID-19. Esta caída se correlaciona con picos de entropía, que indican un periodo de alta incertidumbre en el mercado.

Posteriormente, el precio muestra una recuperación gradual con fluctuaciones, alcanzando un pico en el 2022, seguido de otra caída antes de mostrar una nueva tendencia al alza hacia finales del 2024. Por otra parte, la entropía Esp presenta picos altos en el 2020 y principios del 2021, pero tiende a estabilizarse con fluctuaciones menos pronunciadas; mientras que la entropía Engl muestra un patrón más volátil, sugiriendo periodos alternados de alta y baja incertidumbre en el mercado inglés.

Desde una perspectiva más amplia de la figura 11, el índice de precios al productor (IPP) refleja una tendencia alcista continua desde mediados del 2021 hasta el 2024, lo que indica un aumento en los costos de producción que podrían afectar los márgenes de beneficio de Ecopetrol.

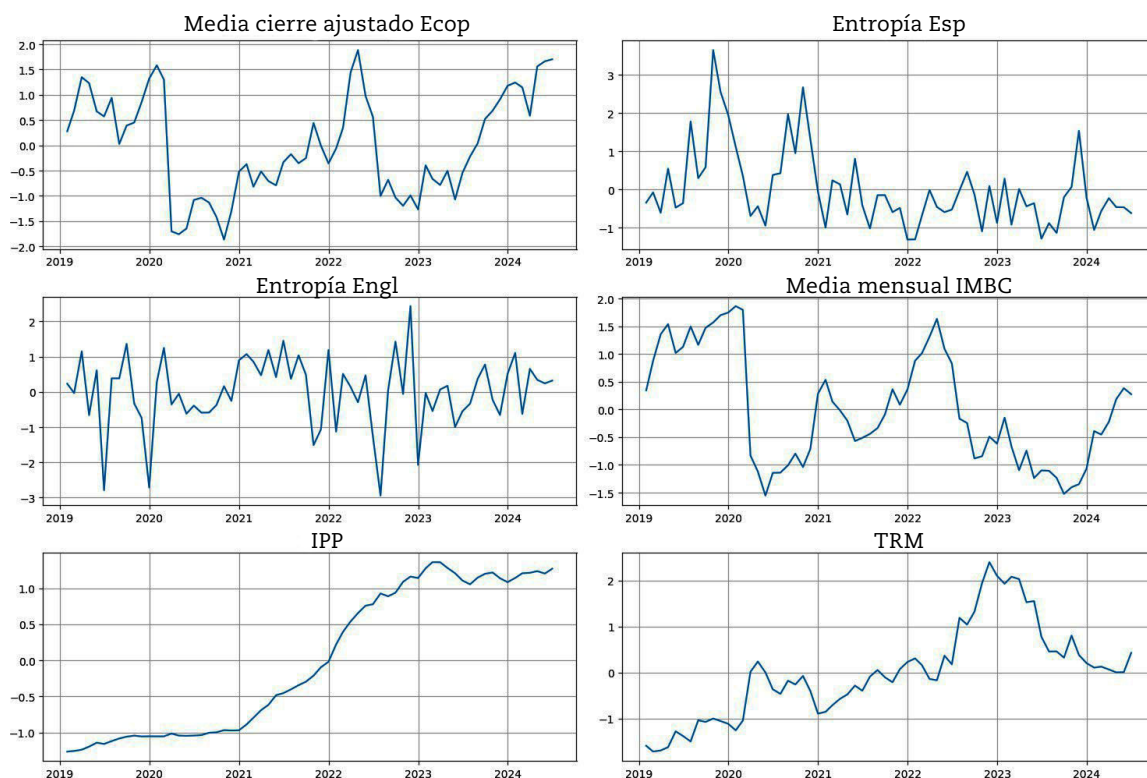


Figura 11. Cierre ajustado Ecopetrol, entropía titulares español, entropía titulares inglés, media mensual del COLCAP, índice de precios al productor, tasa representativa del mercado

Fuente: elaboración propia.

A su vez, la tasa representativa del mercado (TRM) muestra un aumento significativo desde el 2021 hasta principios del 2023, seguido de una corrección a la baja, sugiriendo una depreciación de la moneda local frente al dólar. Esto podría haber beneficiado a Ecopetrol en términos de ingresos por exportaciones. Además, el índice mensual de bienestar del consumidor (IMBC) muestra una relación inversa con el precio de Ecopetrol en ciertos periodos, lo que indica su posible naturaleza como indicador adelantado o rezagado. En conjunto, estos factores pintan un panorama económico complejo y volátil que influye en el desempeño de Ecopetrol y su percepción de riesgo en diferentes mercados.

Los hallazgos sugieren que los inversores deberían monitorear no solo los indicadores macroeconómicos, sino también el sentimiento del mercado desde diferentes titulares (ver figura 12), ya que este último tiene un impacto significativo en la volatilidad de los activos de Ecopetrol. Incluso, se evidencia una correlación negativa entre el aumento en la percepción del sentimiento positivo y la volatilidad en el precio de las acciones.

Conclusiones

El análisis realizado evidencia la relevancia de las métricas de sentimiento, entropía y volatilidad para comprender los movimientos del mercado financiero, específicamente en el contexto de Ecopetrol. La integración de modelos avan-

zados de inteligencia artificial de procesamiento de lenguaje natural, como FinBERT y XLM-RoBERTa, permite capturar patrones complejos en datos textuales multilingües, destacando su capacidad para identificar señales de incertidumbre y comportamientos de mercado.



Figura 12. Nube de palabras de titulares

Fuente: elaboración propia.

Los resultados muestran una fuerte correlación entre los picos de entropía en titulares, tanto en inglés como en español, y eventos significativos en el precio de la acción, lo que refuerza la importancia de monitorear las narrativas mediáticas como indicadores adelantados de inestabilidad o recuperación.

Asimismo, el uso de componentes principales y su relación con variables como persistencia e impacto ofrece un marco robusto para simplificar la interpretación de datos, en

el que se destaca la interacción entre duración e intensidad de los fenómenos analizados.

En términos gerenciales y estratégicos, este estudio subraya la necesidad de integrar herramientas cuantitativas y cualitativas en la toma de decisiones financieras. La convergencia de indicadores, como la volatilidad GARCH, la entropía y el sentimiento, proporciona un enfoque integral para evaluar riesgos y oportunidades en el mercado.

Además, el análisis de factores macroeconómicos, como el índice de precios al productor (IPP) y la tasa representativa del mercado (TRM), sugiere un entorno dinámico en el que interactúan variables globales y locales para influir en el desempeño de Ecopetrol. Este panorama, enriquecido con datos multivariados y su respectiva interpretación, permite a los responsables de la toma de decisiones anticipar periodos de inestabilidad, adaptar estrategias operativas y optimizar carteras de inversión con base en un entendimiento más profundo de los factores que afectan al mercado energético.

A manera de recomendación, las directivas de Ecopetrol podrían incorporar el análisis de sentimientos en sus estrategias de comunicación, ya que esto puede influir positivamente en la percepción del mercado. Para los inversores, se sugiere utilizar herramientas de análisis de sentimientos como parte de su proceso de toma de decisiones.

Los resultados de este estudio son consistentes con las tendencias globales que destacan la importancia de los factores no financieros, como el sentimiento del mercado, en la volatilidad de los activos. El sentimiento del mercado influye en el comportamiento de las acciones en mercados emergentes.

Por último, es importante anotar que este estudio tiene algunas limitaciones, como la dependencia de datos históricos. Futuras investigaciones podrían explorar el uso de modelos más avanzados o la inclusión de datos en tiempo real. Además, se recomienda ampliar el análisis a otras empresas del sector energético para validar los hallazgos.

Financiación

Este trabajo no ha recibido financiación alguna para su realización y su objetivo final es contribuir al avance del conocimiento en este campo.

Conflicto de interés

Los autores declaran no tener conflicto de intereses en la elaboración de este documento. El contenido y las opiniones expresadas en él no comprometen a las instituciones referenciadas en el perfil de los autores.

Contribución de los autores

Fernando Cárdenas García: conceptualización, redacción de revisión y edición, análisis formal, escritura (correcciones del arbitraje y de edición); Gregorio Izquierdo Llanes:

análisis formal, curación de datos, escritura (borrador original), investigación, metodología; José Gabriel Orosco: conceptualización, investigación, metodología, software.

Referencias

- Atif, M., Raza Rabbani, M., Bawazir, H., Hawaldar, I. T., Chebab, D., Karim, S., & AlAbbas, A. (2022). Oil price changes and stock returns: Fresh evidence from oil exporting and oil importing countries. *Cogent Economics & Finance*, 10(1), 2018163. <https://doi.org/10.1080/23322039.2021.2018163>
- Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593-1636. <https://doi.org/10.1093/qje/qjw024>
- Baker, S. R., Bloom, N., Davis, S. J., & Kost, K. J. (2019). Policy news and stock market volatility (NBER Working Paper No. 25720). National Bureau of Economic Research. <https://doi.org/10.3386/w25720>
- Bermpei, T., Kalyvas, A. N., Neri, L., & Russo, A. (2022). Does economic policy uncertainty matter for financial reporting quality? Evidence from the United States. *Review of Quantitative Finance and Accounting*, 58(2), 795-845. <https://doi.org/10.1007/s11156-021-01010-2>
- Biktimirov, E. N., Sokolyk, T., & Ayanso, A. (2021). Sentiment and hype of business media topics and stock market returns during the COVID-19 pandemic. *Journal of Behavioral and Experimental Finance*, 31, 100542. <https://doi.org/10.1016/j.jbef.2021.100542>
- Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3), 623-685. <https://doi.org/10.3982/ECTA6248>
- Candelo Viáfara, J. M., Oviedo Gómez, A., & Lozano Mejía, E. (2023). Macroeconomía y mercado bursátil: el impacto y la transmisión de volatilidad de las variables macroeconómicas al mercado bursátil colombiano. *Revista Facultad de Ciencias Económicas*, 31(1), 103-117. <https://doi.org/10.18359/rfce.6413>
- Chauhan, Y., & Jaiswall, M. (2023). Economic policy uncertainty and incentive to smooth earnings. *International Review of Economics & Finance*, 85, 93-106. <https://doi.org/10.1016/j.iref.2023.01.014>
- Chen, P., Boukouvalas, Z., & Corizzo, R. (2024). A deep fusion model for stock market prediction with news headlines and time series data. *Neural Computing & Applications*, 36, 21229-21271. <https://doi.org/10.1007/s00521-024-10303-1>
- Chen, X., Xie, H., Li, Z., Zhang, H., Tao, X., & Wang, F. L. (2025). Sentiment analysis for stock market research: A bibliometric study. *Natural Language Processing Journal*, 10, 100125. <https://doi.org/10.1016/j.nlp.2025.100125>
- Chen, Y., Dong, S., Qian, S., & Chung, K. (2024). Impact of oil price volatility and economic policy uncertainty on business investment-Insights from the energy sector. *Heliyon*, 10(5), e26533. <https://doi.org/10.1016/j.heliyon.2024.e26533>
- Choi, S., & Yoon, C. (2022). Uncertainty, financial markets, and monetary policy over the last century. *The BE Journal of Macroeconomics*, 22(2), 397-434. <https://doi.org/10.1515/bejm-2020-0013>
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., & Lample, G. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440-8451. <https://doi.org/10.48550/arXiv.1911.02116>
- Das, N., Sadhukhan, B., Chatterjee, R., & Chakrabarti, S. (2024a). Integrating sentiment analysis with graph neural networks for enhanced stock prediction: A comprehensive survey. *Decision Analytics Journal*, Article 100417. <https://doi.org/10.1016/j.dajour.2024.100417>
- Das, N., Sadhukhan, B., Bhakta, S.S., Chakrabarti, S. (2024b). Integrating EEMD and ensemble CNN with X (Twitter) sentiment for enhanced stock price predictions. *Social Network Analysis and Mining*, 14(1), Article 29. <https://doi.org/10.1007/s13278-023-01190-w>

- Delgadillo, J., Kinyua, J., & Mutigwe, C. (2024). FinSoSent: Advancing financial market sentiment analysis through pretrained large language models. *Big Data and Cognitive Computing*, 8(8), 87. <https://doi.org/10.3390/bdcc8080087>
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Volume 1 (Long and Short Papers), (pp. 4171-4186). Association for Computational Linguistics. <https://aclanthology.org/N19-1423/>
- Du, K., Xing, F., Mao, R., & Cambria, E. (2024). Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys*, 56(1), 1-42. <https://doi.org/10.1145/3649451>
- Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters*, 9(3), 167-175. <https://doi.org/10.1016/j.frl.2011.10.003>
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. <https://doi.org/10.2307/2325486>
- Fatourous, G., Soldatos, J., Kouroumalis, K., Makridakis, G., & Kyriazis, D. (2023). Transforming sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications*, 14, 100508. <https://doi.org/10.1016/j.mlwa.2023.100508>
- Hao, J., He, F., Ma, F., Zhang, S., & Zhang, X. (2025). Machine learning vs deep learning in stock market investment: An international evidence. *Annals of Operations Research*, 348(1), Article 105654, 93-115. <https://doi.org/10.1007/s10479-023-05286-6>
- Heydarian, P., Bifet, A., & Corbet, S. (2025). Understanding market sentiment analysis: A survey. *Journal of Economic Surveys*, 39(3), 1125-1147. <https://doi.org/10.1111/joes.12645>
- Inserte, P. R., Nakhle, M., Qader, R., Caillaut, G., & Liu, J. (2024). Large language model adaptation for financial sentiment analysis. *arXiv preprint arXiv:2401.14777*. <https://doi.org/10.48550/arXiv.2401.14777>
- Kellogg, R. (2014). The effect of uncertainty on investment: Evidence from Texas oil drilling. *American Economic Review*, 104(6), 1698-1734. <https://doi.org/10.1257/aer.104.6.1698>
- Kjærland, F., Kosberg, F., & Misje, M. (2021). Accrual earnings management in response to an oil price shock. *Journal of Commodity Markets*, 22, 100138. <https://doi.org/10.1016/j.jcomm.2020.100138>
- Liapis, C. M., Karanikola, A., & Kotsiantis, S. (2023). Investigating deep stock market forecasting with sentiment analysis. *Entropy*, 25(2), 219. <https://doi.org/10.3390/e25020219>
- Liu, C., Arulappan, A., Naha, R., Mahanti, A., Kamruzzaman, J., & Ra, I. H. (2024). Large language models and sentiment analysis in financial markets: A review, datasets and case study. *Ieee Access*. <https://doi.org/10.1109/ACCESS.2024.3445413>
- Loginova, E., Tsang, W. K., van Heijningen, G., Kerkhove, L. P., & Benoit, D. F. (2024). Forecasting directional bitcoin price returns using aspect-based sentiment analysis online text data. *Machine Learning*, 113(7), 4761-4784. <https://doi.org/10.1007/s10994-021-06095-3>
- Mamman, S. O., Iliyasu, J., Ahmed, U. A., & Salami, F. (2024). Global uncertainties, geopolitical risks and price exuberance: Evidence from international energy market. *OPEC Energy Review*, 48(2), 63-120. <https://doi.org/10.1111/opec.12297>
- Nannepagu, M., Babu, D. B., & Madhuri, C. B. (2024). Leveraging hybrid AI models: DQN, Prophet, BERT, ART-NN, and transformer-based approaches for advanced stock market forecasting. *International Journal of Engineering Science and Advanced Technology*, 24(12), 144-151. https://www.ijesat.com/ijesat/files/V24I1220_1735359632.pdf
- Panousi, V., & Papanikolaou, D. (2012). Investment, idiosyncratic risk, and ownership. *The Journal of Finance*, 67(3), 1113-1148. <https://doi.org/10.1111/j.1540-6261.2012.01743.x>
- Paule Vianez, J. (2020). *Influencia de la incertidumbre de política económica en los mercados financieros* [tesis doctoral, Universidad Rey Juan Carlos]. <http://hdl.handle.net/10115/17553>
- Rao, A., Tedeschi, M., Mohammed, K. S., & Shahzad, U. (2024). Role of economic policy uncertainty in energy commodities prices forecasting: Evidence from a hybrid deep learning approach. *Computational Economics*, 64(6), 3295-3315. <https://doi.org/10.1007/s10614-024-10550-3>
- Rigamonti, A. P., Greco, G., Pierotti, M., & Capocchi, A. (2024). Macroeconomic uncertainty and earnings management: Evidence from commodity firms. *Review of Quantitative Finance and Accounting*, 62(4), 1615-1649. <https://doi.org/10.1007/s11156-024-01246-8>
- Sáenz del Río, D. (2016). *El impacto de los precios del petróleo y el dólar en los estados financieros de Ecopetrol S. A.* [tesis de pregrado, Universidad de La Sabana]. <http://hdl.handle.net/10818/21908>
- Sayeed, M. S., Mohan, V., & Sonai Muthu Anbananthen, K. (2023). Bert: A review of applications in sentiment analysis. *HighTech and Innovation Journal*, 4(2), 453-462. <https://doi.org/10.28991/HIJ-2023-04-02-015>
- Shobayo, O., Adeyemi-Longe, S., Popoola, O., & Ogunleye, B. (2024). Innovative sentiment analysis and prediction of stock price using FinBERT, GPT-4 and logistic regression: A data-driven approach. *Big Data and Cognitive Computing*, 8(11), 143. <https://doi.org/10.3390/bdcc8110143>
- Siddique, M. T., Jamee, S. S., Sajal, A., Mou, S. N., Mahin, M. R. H., Obaid, M. O., ..., & Hasan, M. (2025). Enhancing automated trading with sentiment analysis: Leveraging large language models for stock market predictions. *The American Journal of Engineering and Technology*, 7(03), 185-195. <https://doi.org/10.37547/tajet/Volume07Issue03-16>
- Smales, L. A. (2021). Geopolitical risk and volatility spillovers in oil and stock markets. *Quarterly Review of Economics and Finance*, 80, 358-366. <https://doi.org/10.1016/j.qref.2021.03.008>
- Szczygielski, J. J., Charteris, A., Bwanya, P. R., & Brzeszczyński, J. (2024). Google search trends and stock markets: Sentiment, attention or uncertainty? *International Review of Financial Analysis*, 91, 102549. <https://doi.org/10.1016/j.irfa.2023.102549>
- Tedeschi, M., Foglia, M., Bouri, E., & Dai, P. F. (2024). How does climate policy uncertainty affect financial markets? Evidence from Europe. *Economics Letters*, 234, 111443. <https://doi.org/10.1016/j.econlet.2023.111443>
- Trust, P., Zahran, A., & Minghim, R. (2023). Understanding the influence of news on society decision making: Application to economic policy uncertainty. *Neural Computing and Applications*, 35(20), 14929-14945. <https://doi.org/10.1007/s00521-023-08438-8>
- Yin, H., & Yang, Q. (2023). Investor sentiment mining based on Bi-LSTM model and its impact on stock price bubbles. *Studies in Nonlinear Dynamics & Econometrics*, 27(1), 1-25. <https://doi.org/10.1515/snde-2022-0028>
- Zhao, L., Li, L., Zheng, X., & Zhang, J. (2021). A BERT based sentiment analysis and key entity detection approach for online financial texts. In *2021 IEEE 24th International conference on computer supported cooperative work in design (CSCWD)* (pp. 1233-1238). IEEE.