

CLASIFICACIÓN AUTOMÁTICA DE NOTICIAS DEL MERCADO DEL COBRE Y SU CORRESPONDENCIA CON VARIACIONES REALES DEL PRECIO DEL MINERAL

FRANCISCO GRANT PALZA

fgrant@fen.uchile.cl

FRANCIS SALDÍAS SANHUEZA

fsaldias@fen.uchile.cl

ABSTRACT

Este estudio tiene como objetivo desarrollar un modelo de Procesamiento de Lenguaje Natural (NLP) capaz de clasificar noticias históricas relacionadas con el mercado del cobre según su impacto potencial (positivo, negativo o neutro) en el precio del mineral, y evaluar si estas clasificaciones coinciden con las variaciones diarias observadas en el precio del cobre.

Se recopilieron aproximadamente 1000 noticias provenientes de diversas fuentes internacionales desde el 2023 a noviembre de 2025. De forma paralela, se obtuvo la serie de precios diarios del cobre desde mayo de 2022 a la fecha. Los resultados muestran una exactitud global del 40%, con un desempeño superior en la identificación de señales positivas, donde el modelo alcanzó una precisión del 56% y un recall del 60%. En contraste, el rendimiento en las categorías negativa y neutra es considerablemente menor, lo que se explica en parte por su menor frecuencia y la mayor complejidad semántica de estos casos.

En consecuencia, su capacidad explicativa respecto de las variaciones reales del precio del cobre sigue siendo limitada. Se discuten implicancias y posibles extensiones hacia modelos explicativos y predictivos más robustos, incluyendo la ampliación del conjunto de entrenamiento, el balanceo de clases, el uso de modelos financieros especializados, o bien, la aplicación a rubros no financieros.

INTRODUCCIÓN

El precio del cobre es un indicador macroeconómico de alta relevancia para la economía mundial y, en particular, para Chile, dado su rol como principal productor global de cobre. Las variaciones en el precio de dicho mineral responden a múltiples factores, entre ellos shocks de oferta, disrupciones operativas, cambios en inventarios, fluctuaciones en la demanda global y decisiones de los principales compradores, especialmente China y Estados Unidos. Dada su importancia en la transición energética, la electromovilidad y la innovación tecnológica, comprender los determinantes del precio del cobre constituye un desafío central para analistas, inversionistas y autoridades económicas.

En mercados altamente sensibles a la información, las noticias juegan un rol clave en la formación de expectativas. Surge así la interrogante respecto de si las noticias relacionadas con el mercado del cobre contienen señales capaces de anticipar o reflejar movimientos del precio. En este

contexto, las técnicas de Procesamiento de Lenguaje Natural (NLP) ofrecen una oportunidad para automatizar la extracción de información desde grandes volúmenes de texto para identificar patrones semánticos.

En base a lo anterior, el presente estudio desarrolla un modelo basado en Llama 3.1, configurado para clasificar cada noticia según su impacto potencial sobre el precio del cobre, ya sea positivo, negativo o neutro. Posteriormente, se evalúa si estas clasificaciones se alinean con las variaciones reales del precio del mineral, lo cual permite medir si el modelo captura adecuadamente señales informativas relevantes para el comportamiento del mercado y si las noticias realmente influyen en el precio del mineral.

A partir de este marco, el estudio se guía por las siguientes preguntas de investigación:

1. ¿Es capaz un modelo NLP de clasificar correctamente el efecto potencial de noticias del mercado del cobre?

2. ¿Coinciden estas clasificaciones con las variaciones reales del precio del mineral?
3. ¿Presenta el modelo un desempeño diferenciado según el tipo de efecto (positivo, negativo o neutro)?
4. ¿Puede un modelo de clasificación de noticias servir como base para futuros modelos predictivos del precio del cobre?

Finalmente, se discuten las implicancias de los resultados para el análisis de riesgo, el monitoreo de mercados y el desarrollo de modelos predictivos que integren información textual y series temporales financieras.

METODOLOGÍA

El estudio se desarrolló en cinco etapas principales: recopilación y preprocesamiento de noticias, clasificación automática mediante un modelo NLP, preparación y normalización de variables, integración con precios históricos del cobre y evaluación del desempeño explicativo del modelo.

Noticias del mercado del cobre

En primera instancia se obtuvo una base de noticias mediante un proceso de extracción y recolección automática desde medios especializados en la industria minera y financiera, tales como Mining y Google News y Yahoo Finance. Este procedimiento realizado, en mayor medida, mediante código python permitió reunir más de 1000 noticias publicadas entre 2023 y 2025, a las que se aplicó un procedimiento de limpieza y de recolección de información. El primer procedimiento consistió en la obtención de la fecha, verificar la validez de la publicación y la remoción de etiquetas HTML. Mientras que el segundo permitió obtener el titular de la noticia y un resumen de esta, quedando un archivo con las variables Fecha, Link, Titular y resumen de la noticia. Esta base output se ocupó, posteriormente, en el análisis Llama para la clasificación del efecto de la noticia en el precio del cobre.

Modelo NLP

La clasificación de noticias se realizó mediante el modelo Llama 3.1 con el que se diseñó un prompt de tipo role-play que permite al

modelo ser un analista económico en la industria minera para facilitar el análisis de la variedad de noticias encontradas. Con esto, el modelo puede definir si la publicación tiene un efecto positivo, negativo o neutro en el precio del metal rojo.

La clasificación se fundamenta en elementos asociados a los determinantes clásicos de oferta y demanda, tales como interrupciones de suministro, variaciones en inventarios, anuncios de expansión de capacidad, estímulos económicos o señales de desaceleración industrial. El modelo produce una salida estructurada en formato: “Etiqueta: Positiva/Negativa/Neutra | Razon: explicación breve”.

Posteriormente, estas respuestas fueron transformadas a numéricas con posibles valores enteros $\{-1,1\}$, asignando +1 a noticias positivas, -1 a negativas y 0 a noticias clasificadas como neutras.

Serie Histórica del precio del cobre



Gráfica N°1: Precio Histórico de Futuro del Cobre

La serie se obtuvo mediante el uso de la web en páginas especializadas en finanzas, decidiendo utilizar Investing View que entrega el Precio Futuro del Cobre de manera diaria al cierre desde mayo del 2022 hasta el 3 de diciembre de 2025. Para cada fecha se calculó la variación porcentual diaria (*Change*). Además, se derivó la variación del día siguiente (*Change_next*), la cual permite evaluar cómo se comportó efectivamente el mercado después de la publicación de cada noticia.

Integración de bases

La integración entre noticias y precios se realizó alineando temporalmente cada artículo con el valor del cobre correspondiente al día de

su publicación. Dado que algunas noticias obtenidas desde la automatización no contenían la fecha de publicación, la base se redujo a 760 noticias. Además, gracias a la normalización numérica aplicada a la Etiqueta, se calculó un promedio ponderado de las etiquetas de las noticias publicadas un mismo día, para así obtener la predicción esperado al día siguiente de la publicación de una noticia, siendo posible obtener una predicción alcista, bajista o neutra. Con este ajuste, la base se redujo a 166 artículos.

A partir de la variación del día siguiente, se construyó la variable objetivo denominada *efecto_real*, definida como positiva si el cambio supera +0.3%, negativa si cae bajo -0.3% y neutra cuando oscila dentro de dicho rango. Esta clasificación permitió comparar directamente la etiqueta generada por el modelo NLP con el comportamiento real del precio del cobre, evaluando su desempeño mediante métricas estándar de clasificación: matriz de confusión, precisión, recall, F1-score y accuracy global.

RESULTADOS

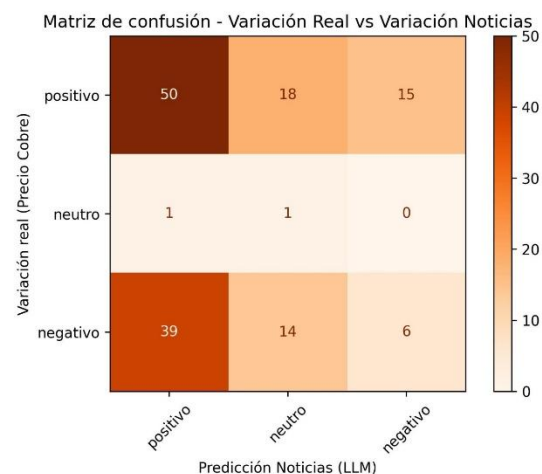
El modelo NLP logró clasificar todas las noticias del conjunto analizado en efectos positivos, negativos o neutros sobre el precio del cobre, utilizando reglas económicas explícitas vinculadas a oferta, demanda, inventarios y disrupciones operativas. Al evaluar si estas clasificaciones se alinean con la variación real del precio del mineral al día siguiente, se observa que el desempeño es heterogéneo entre clases. La mayoría de las predicciones del modelo se concentran en la categoría positiva, lo cual coincide parcialmente con la distribución real de días con alzas del precio. Sin embargo, la capacidad para identificar efectos negativos o neutros es considerablemente menor.

Los resultados cuantitativos muestran esta asimetría. La clase positiva obtiene una precisión de 0,56 y un recall de 0,60, lo que sugiere que el modelo captura una fracción relevante de las señales alcistas presentes en las noticias. Por el contrario, la clase negativa presenta un recall de 0,10 y la clase neutra, pese a tener solo dos casos, un recall de 0,50, aunque con precisión cercana a cero. En consecuencia, la exactitud global del modelo

alcanza un 40%, indicando que, si bien existe una capacidad parcial para identificar noticias asociadas a aumentos en el precio del cobre, las señales bajistas y neutras continúan siendo difíciles de extraer desde el texto.

La matriz de confusión lo refleja claramente:

- De 83 días realmente positivos, el modelo acierta 50.
- De 59 días realmente negativos, solo identifica 6.
- La clase neutra aparece con muy baja frecuencia, lo que complica su aprendizaje y evaluación.



Gráfica N°2: Mapa de Calor de la Matriz de Confusión

Clase	Precisión	Recall	F1-score	SopORTE
Positivo	0,56	0,60	0,58	83
Neutro	0,03	0,50	0,06	2
Negativo	0,29	0,10	0,15	59
Accuracy			0,40	144
Macro Average	0,29	0,40	0,26	144

Weighted Average	0,44	0,40	0,40	144
------------------	------	------	------	-----

Tabla N°1: Métricas

DISCUSIÓN

Los resultados obtenidos permiten sostener que el modelo NLP empleado fue capaz de identificar señales direccionales relevantes en las noticias del mercado del cobre; sin embargo, su desempeño general evidencia limitaciones metodológicas y de capacidad clasificatoria que deben ser analizadas en profundidad.

El desempeño deficiente en las clases negativa y neutra revela limitaciones relevantes. Por un lado, los factores que presionan a la baja el precio del cobre suelen estar vinculados a elementos macroeconómicos más difusos (como desaceleraciones graduales, señales de debilidad manufacturera o fortalecimiento del dólar), cuya textualidad es más ambigua y difícil de capturar sin un modelo entrenado específicamente en finanzas. Por otro lado, la escasa presencia de días realmente neutros produce un desbalance estructural que dificulta el aprendizaje de esta categoría. Además, el uso del retorno del día siguiente ($t+1$) como referencia introduce ruido, ya que el precio del cobre responde a múltiples factores exógenos no contenidos en las noticias.

Estos elementos sugieren que, aunque la aproximación NLP capta parte del contenido económico relevante, su capacidad clasificatoria sigue siendo limitada en mercados altamente volátiles y multifactoriales como el cobre.

CONCLUSIONES

En su configuración actual, el modelo NLP no constituye una base predictiva robusta para anticipar movimientos del precio del cobre. A pesar de su capacidad para identificar una proporción significativa de días con variaciones positivas, su desempeño se deteriora al clasificar efectos negativos y neutros, lo cual reduce su utilidad práctica para modelamiento predictivo. Estas limitaciones provienen tanto del desbalance de clases como de la complejidad estructural del propio mercado del cobre, en el cual

factores globales, financieros y geopolíticos influyen simultáneamente en la formación de precios.

No obstante, los resultados son valiosos como punto de partida. El análisis muestra que las noticias contienen señales informativas, pero su extracción requiere modelos especializados en lenguaje financiero o modelos supervisados entrenados sobre un conjunto de noticias etiquetadas manualmente. Asimismo, futuros estudios podrían explorar horizontes temporales más amplios, integrar variables macroeconómicas o combinar NLP con modelos de series temporales. Más allá del cobre, la metodología podría aplicarse en mercados donde los efectos noticiosos son más directos o donde la volatilidad es menor, ampliando el potencial del enfoque a sectores industriales, políticos o sociales.

Bibliografía

- Bollen, J., Mao, H., & Zeng, X. (2010). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Cochilco. (2025, 2 de noviembre). *Inicio: Comisión Chilena del Cobre*. Comisión Chilena del Cobre. <https://www.cochilco.cl/web/>
- Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Tetlock, P. (2017). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3), 1109–1148. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Roziere, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). *LLaMA: Open and Efficient Foundation Language Models* (Preimpresión). arXiv. <https://doi.org/10.48550/arXiv.2302.13971>
- Zhang, X., Fuehres, H., & Gloor, P. (2011). Predicting Stock Market Indicators Through

Twitter “I hope it is not as bad as I fear”.
Procedia - Social and Behavioral Sciences, 26,

154–162.

<https://doi.org/10.1016/j.sbspro.2011.10.562>