**Título:**

Integración del análisis de sentimiento noticioso en modelos predictivos del precio del café arábica: un caso de su aplicación en la gestión de riesgos para la industria cafetera colombiana desde la perspectiva de compra.

**Autor:**

Diego Parra

**Afiliación:**

Estudiante de Maestría en Ciencias de Datos, Escuela Colombiana de Ingeniería Julio Garavito

**Correo Electrónico:**

diego.parra-v@mail.escuelaing.edu.co2. Resumen y palabras clave

**Resumen:**

Este estudio evalúa la integración de modelos de predicción del precio del café arábica con el análisis de sentimiento de noticias. Empleando modelos híbridos que combinan técnicas de machine learning para analizar los precios históricos del commodity adicionando una variable de procesamiento de lenguaje natural, que extrae el sentimiento de noticias relacionadas con el mercado del café como el clima en los principales países productores, eventos geopolíticos, y fluctuaciones en la demanda.

El objetivo es demostrar que la inclusión del sentimiento noticioso mejora la capacidad predictiva de los modelos tradicionales. El resultado estructura la base de un modelo de gestión de riesgos para la industria colombiana de café mitigando la volatilidad del mercado y permitiendo una toma de decisiones más informada.

**Palabras claves:**

Análisis de Sentimiento, Café Arábica, Gestión de Riesgos, Procesamiento de Lenguaje Natural, shocks , Machine Learning y Series de Tiempo

**Introducción**

La volatilidad del precio del café arábica es un problema persistente y crítico con implicaciones significativas para las economías que dependen de la agricultura, como la de Colombia. Esta volatilidad ha generado un escenario complejo: aunque puede beneficiar a los caficultores en ciertos ciclos de precios altos, a menudo afecta negativamente a otros actores de la cadena, como los comercializadores y, especialmente a los consumidores finales, quienes enfrentan costos elevados por el producto (Manogna, Dharmaji, & Sarang, 2025). Las fluctuaciones de precios pueden desestabilizar los ingresos de los productores, en particular los pequeños agricultores, y desalentar la inversión en el sector (Manogna, Dharmaji, & Sarang, 2025).

Los precios del café están influenciados por una confluencia de factores impredecibles, estos incluyen factores climáticos como sequías o heladas, que pueden causar estragos en las cosechas (Júnior, Júnior, & Zambalde, 2016) (Silva & Ferreira, 2016); cambios en políticas comerciales, como la imposición de aranceles o prohibiciones de exportación (Manogna, Dharmaji, & Sarang, 2025); y conflictos geopolíticos, que abarcan desde tensiones entre gobiernos hasta actos de terrorismo (Özdemir, Vurur, Ozen, Swiecka, & Grima, 2025); por ejemplo la principal fuente de volatilidad en el café arábica se atribuye a las incertidumbres específicas de la producción en Brasil, el mayor productor mundial (Özdemir, Vurur, Ozen, Swiecka, & Grima, 2025).

El mercado es particularmente sensible a las noticias; tanto las "buenas" como las "malas" impactan el precio de manera asimétrica (Özdemir, Vurur, Ozen, Swiecka, & Grima, 2025) (Júnior, Júnior, & Zambalde, 2016). Sin embargo, las noticias negativas, como fallas en las cosechas o condiciones climáticas adversas, tienden a poner más nervioso al mercado, afectando directamente la producción y contribuyendo de forma expresiva a la volatilidad del producto (Júnior, Júnior, & Zambalde, 2016) (Silva & Ferreira, 2016).

Es interesante como un rasgo característico del mercado del café es la persistencia de la volatilidad, donde los shocks en los precios ya sean positivos o negativos, tienen efectos prolongados y tardan un tiempo considerable en disiparse (Silva & Ferreira, 2016). Este fenómeno está estrechamente ligado al efecto apalancamiento o asimetría, observado consistentemente en diversos estudios (Silva & Ferreira, 2016) (Manogna, Dharmaji, & Sarang, 2025) que demuestra que los shocks negativos aumentan la volatilidad en mayor medida que los shocks positivos, Además, es importante señalar que una subida grande y prolongada de precios no siempre es beneficiosa a largo plazo; puede incentivar a los agricultores a sobre producir, lo que a su vez provoca una caída de los precios en ciclos posteriores y desestabiliza aún más el mercado. Por todo ello, es fundamental gestionar el riesgo ante los shocks para garantizar la estabilidad (Silva & Ferreira, 2016).

Por ello, la capacidad de predecir estas fluctuaciones es vital para que todos los actores de la cadena del café, desde productores hasta exportadores, puedan tomar decisiones estratégicas informadas que ayuden a gestionar el riesgo y estabilizar la carga financiera (Özdemir, Vurur, Ozen, Swiecka, & Grima, 2025) (Ngoc, y otros, 2023) (Manogna, Dharmaji, & Sarang, 2025) (Silva & Ferreira, 2016)

Para mejorar la precisión de los pronósticos, es necesario evolucionar más allá de los modelos econométricos tradicionales y adoptar herramientas de aprendizaje automático (ML) (Jin & Xu, 2024) (Manogna, Dharmaji, & Sarang, 2025) (Ngoc, y otros, 2023). Las redes neuronales de memoria a corto y largo plazo (LSTM), por ejemplo, han demostrado ser muy prometedoras para capturar dependencias a largo plazo y patrones no lineales en los datos de precios de las materias primas agrícolas (Manogna, Dharmaji, & Sarang, 2025) (Ngoc, y otros, 2023). Sin embargo, los modelos de aprendizaje automático por sí solos pueden ser insuficientes si no se integran factores externos (Ngoc, y otros, 2023). El análisis de sentimientos extraído de noticias es un factor crucial, ya que los inversionistas y agentes del mercado reaccionan a las "buenas y malas noticias" del día a día, (Júnior, Júnior, & Zambalde, 2016) (Silva & Ferreira, 2016). Un análisis de este tipo puede capturar información cualitativa valiosa sobre el clima, las políticas comerciales, las fluctuaciones de la demanda y los riesgos geopolíticos, los cuales tienen un impacto directo y a menudo asimétrico en la volatilidad del precio del café (Silva & Ferreira, 2016) (Júnior, Júnior, & Zambalde, 2016). De hecho, algunos índices de riesgo geopolítico se construyen directamente a partir del análisis de artículos de prensa (Özdemir, Vurur, Ozen, Swiecka, & Grima, 2025).

Este estudio busca integrar modelos de aprendizaje automático para la predicción de precios con un análisis de sentimientos de noticias mediante el procesamiento de lenguaje natural (PLN) (Ngoc, y otros, 2023). El objetivo es demostrar que esta integración mejora significativamente la capacidad predictiva del precio del café, superando las limitaciones de los modelos independientes (Ngoc, y otros, 2023) (Manogna, Dharmaji, & Sarang, 2025). Con esta base, se busca estructurar un modelo de gestión de riesgos que permita a las empresas de la industria del café arábico colombiano tomar decisiones más informadas, optimizar sus estrategias de cobertura y administrar de manera más efectiva el riesgo financiero inherente a este negocio (Manogna, Dharmaji, & Sarang, 2025), (Silva & Ferreira, 2016) (Júnior, Júnior, & Zambalde, 2016).

**Referencias bibliográficas**

1. Jin, B., & Xu, X. (2024). Machine learning coffee price predictions. *Journal of Uncertain Systems, 17*(4), 2450023. <https://doi.org/10.1142/S1752890924500235>
2. da Silva, C. A. G., & Ferreira, L. da R. (2016). Asymmetric volatility modeling of spot prices of Arabic coffee in Brazil. AgEcon Search. <http://ageconsearch.umn.edu>
3. Lima Júnior, P. O., Castro Júnior, L. G., & Zambalde, A. L. (2016). *Applying textmining to classify news about supply and demand in the coffee market*.
4. Özdemir, L., Vurur, N. S., Ozen, E., Świecka, B., & Grima, S. (2025). Volatility modeling of the impact of geopolitical risk on commodity markets. Economies, 13(4), 88. <https://doi.org/10.3390/economies13040088>
5. Manogna, R. L., Dharmaji, V., & Sarang, S. (2025). A novel hybrid neural network-based volatility forecasting of agricultural commodity prices: empirical evidence from India. Journal of Big Data, 12(85). <https://doi.org/10.1186/s40537-025-01131-8>
6. Le Ngoc, T. N., Lam, D. T., Hai Minh, T. N., Doan, T. C., Nguyen, N. P., Nguyen, H. M., Nguyen, T. N., Tran, L. D., & Tran, N-Q. (2023). Machine learning for agricultural price prediction: A case of coffee commodity in Vietnam market. https://doi.org/10.1109/BCD57833.2023.10466313