

Efecto del Sentimiento en Redes Sociales sobre el Precio del Bitcoin: Un Análisis Empírico

Sebastián Marinovic Leiva¹ Ricardo Iván Lizana Aseña¹
Luis Andrés Gutiérrez González¹

¹Magíster en Data Science, Universidad de Las Américas, Chile

Enero 2026

Resumen

Bitcoin es un activo altamente volátil cuyo comportamiento puede verse influenciado por factores exógenos, entre ellos el sentimiento expresado en redes sociales. Este estudio analiza si el sentimiento extraído de Twitter posee capacidad predictiva sobre variaciones del precio de Bitcoin en ventanas de 1, 6 y 24 horas. Se emplearon técnicas de procesamiento de lenguaje natural (VADER y TextBlob) para cuantificar el sentimiento de 905,863 tweets, integrándolo en modelos predictivos tradicionales (Ridge, Random Forest) y redes neuronales LSTM. Los resultados muestran que, a pesar de encontrar correlación estadísticamente significativa entre sentimiento y retornos ($\rho = 0,088$, $p < 0,01$), la incorporación de variables de sentimiento no mejoró el desempeño predictivo de los modelos (test de Wilcoxon $p = 0,68$). Estos hallazgos sugieren que, para el período analizado (febrero-agosto 2021), el sentimiento social de Twitter no aporta poder predictivo incremental sobre indicadores técnicos tradicionales.

Palabras clave: Bitcoin, análisis de sentimiento, predicción de precios, Twitter, machine learning, LSTM, NLP.

1. Introducción

Bitcoin (BTC) es el activo digital más representativo del mercado de criptomonedas, caracterizado por alta volatilidad y sensibilidad a factores exógenos [1]. A diferencia de los mercados tradicionales, su dinámica puede estar condicionada por señales de atención pública y percepción colectiva. Con el auge de las redes sociales, los mercados financieros reaccionan en tiempo real a información difundida en plataformas como Twitter y Reddit, donde publicaciones virales pueden coincidir con movimientos de precio significativos.

La literatura ha explorado el uso de Twitter como fuente de señales predictivas para mercados financieros. Garcia y Schweitzer [2] reportaron asociaciones entre indicadores sociales y dinámica del mercado de Bitcoin, mientras que Kaminski [3] desarrolló estrategias de trading algorítmico basadas en señales sociales. En paralelo, el análisis de sentimiento es una línea consolidada en procesamiento de lenguaje natural (NLP) para cuantificar el tono emocional de textos [4,5].

Este trabajo integra ambas áreas para evaluar empíricamente la capacidad predictiva del sentimiento social sobre el precio de Bitcoin. La pregunta central de investigación es:

¿Puede el sentimiento expresado en redes sociales como Twitter anticipar movimientos significativos en el precio futuro del Bitcoin, y esta señal mejora la exactitud de

modelos predictivos de series temporales?

2. Marco Teórico

2.1. Volatilidad de Bitcoin y factores exógenos

La volatilidad de Bitcoin se asocia a múltiples factores: condiciones macroeconómicas, liquidez del mercado, regulaciones y señales de atención pública. Kristoufek [1] documentó relación significativa entre búsquedas en Google Trends y Wikipedia con la dinámica de precios de BTC.

2.2. Análisis de sentimiento en finanzas

El análisis de sentimiento busca cuantificar el tono emocional de un texto clasificándolo como positivo, negativo o neutro [4,5]. En el contexto de redes sociales, herramientas léxicas como VADER han demostrado buen desempeño [6]. Enfoques basados en Transformers, como BERT, capturan contexto bidireccional [7].

2.3. Predicción con redes neuronales recurrentes

Las redes LSTM han demostrado efectividad para modelar dependencias de largo plazo en series temporales [8]. Estudios recientes han aplicado LSTM a la predicción de criptomonedas incorporando variables de sentimiento social con resultados mixtos [9,10,11].

3. Metodología

3.1. Datos

Los datos provienen de dos fuentes principales:

- **Tweets:** 1,000,025 publicaciones relacionadas con Bitcoin de Kaggle (febrero-agosto 2021).
- **Precio BTC:** Serie OHLCV con frecuencia horaria de Binance.

Cuadro 1: Estadísticas descriptivas del conjunto de datos

Variable	Valor
Período analizado	05-Feb-2021 a 21-Ago-2021
Total tweets (raw)	1,000,025
Tweets tras limpieza	905,863
Observaciones OHLCV	4,742 horas
Observaciones para modelado	1,156
Cobertura temporal	32.0 %
Promedio tweets/hora	190.6

3.2. Análisis de sentimiento

Se emplearon dos enfoques:

- **VADER** [6]: Léxico para redes sociales con score compound [-1, 1].

- **TextBlob**: Polaridad [-1, 1] y subjetividad [0, 1].

Se crearon indicadores FOMO y FUD mediante conteo de palabras clave.

3.3. Variables objetivo

Se definieron tres horizontes de predicción:

$$r_{t+H} = \ln(P_{t+H}) - \ln(P_t), \quad H \in \{1, 6, 24\} \text{ horas} \quad (1)$$

3.4. Modelos

Se compararon: Naive (Zero Return), Ridge Regression, Random Forest (500 árboles), y LSTM. Cada modelo se evaluó en configuración BASE (19 features técnicas) y BASE+SENT (145 features).

4. Resultados

4.1. Análisis exploratorio

La Figura 1 muestra la evolución del precio de Bitcoin, con máximo cercano a \$65,000 USD en abril 2021.

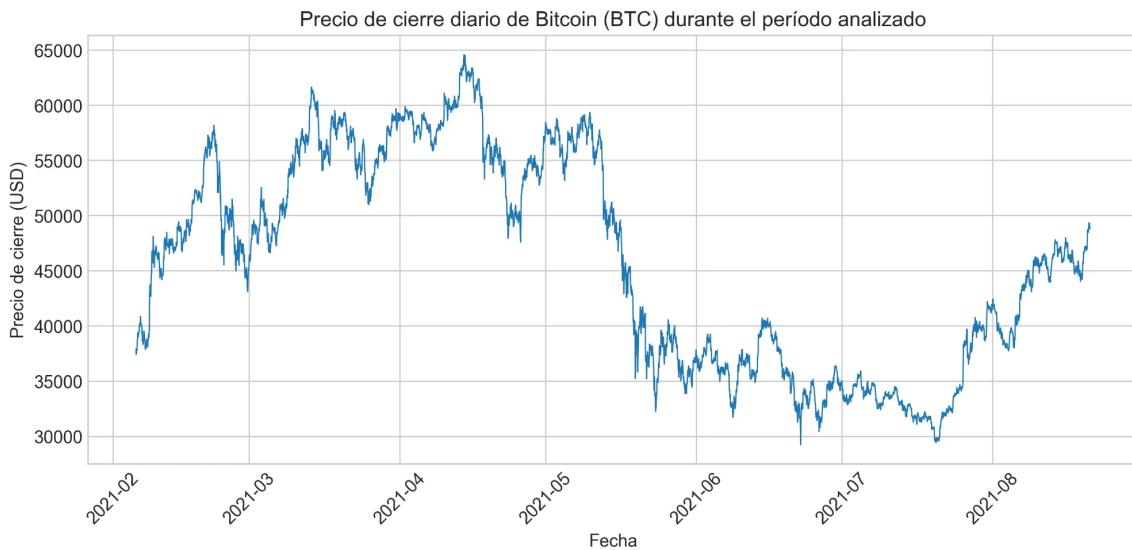


Figura 1: Precio de cierre horario de Bitcoin (BTC) durante el período analizado.

4.2. Distribución de sentimiento

El análisis reveló sesgo positivo (Figura 2): Positivo 52.9 %, Neutral 32.6 %, Negativo 14.5 %.

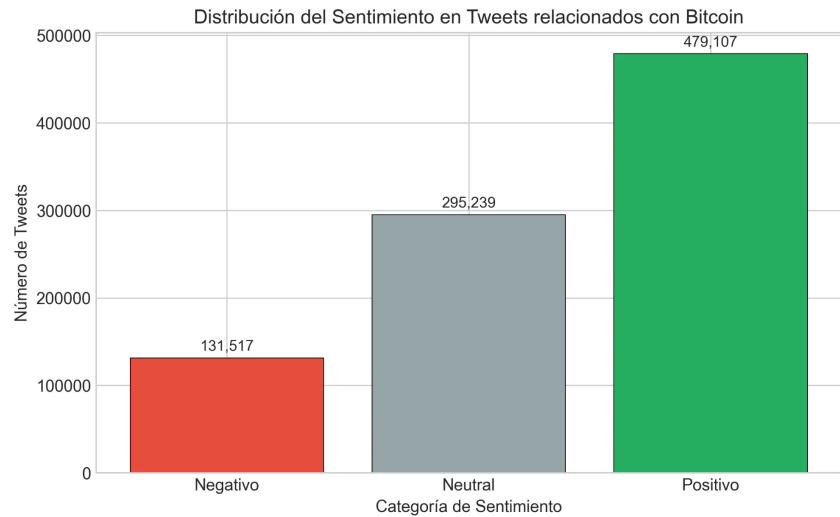


Figura 2: Distribución del sentimiento en tweets relacionados con Bitcoin.

4.3. Relación sentimiento-precio

La Figura 3 muestra que el sentimiento se mantiene positivo independientemente del precio.

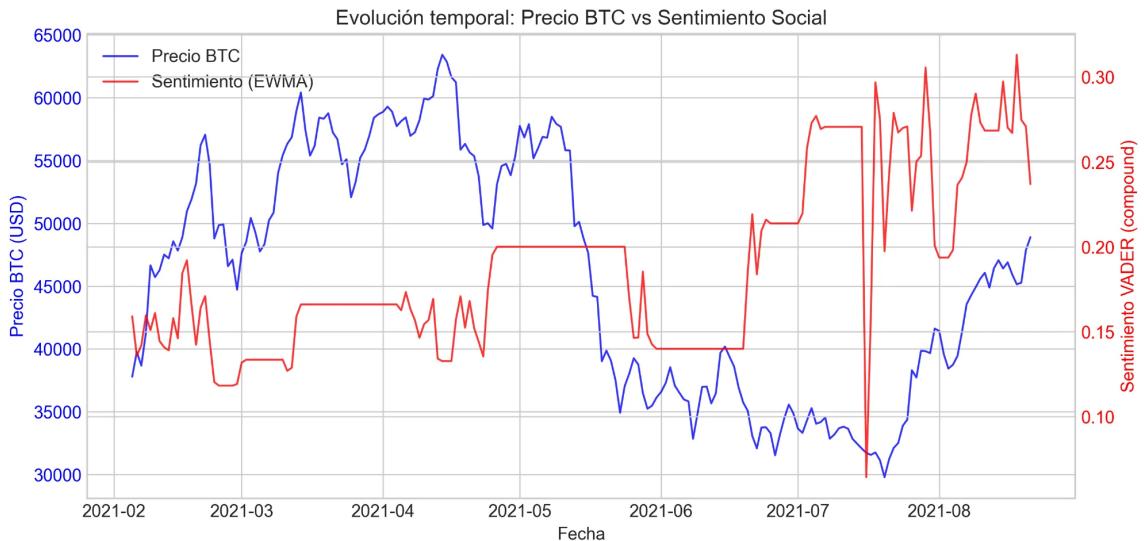


Figura 3: Evolución temporal del precio de Bitcoin y sentimiento social.

4.4. Análisis de correlación

La matriz de correlación (Figura 4) revela correlación negativa entre precio y sentimiento ($r = -0,37$).

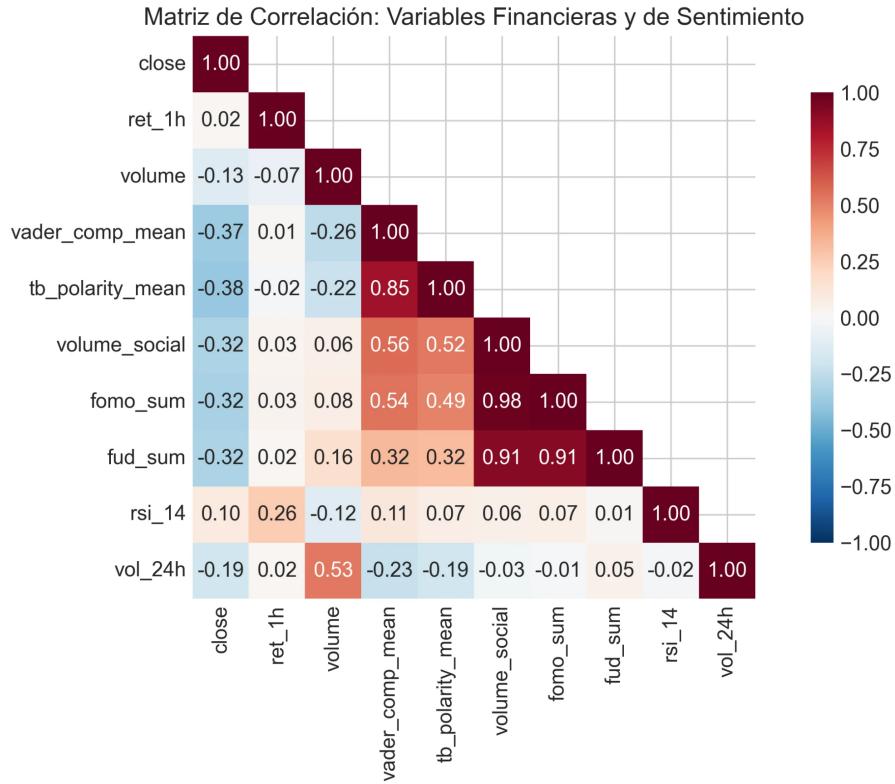


Figura 4: Matriz de correlación entre variables financieras y de sentimiento.

La correlación rolling (Figura 5) muestra relación inestable que oscila entre -0.4 y +0.5.

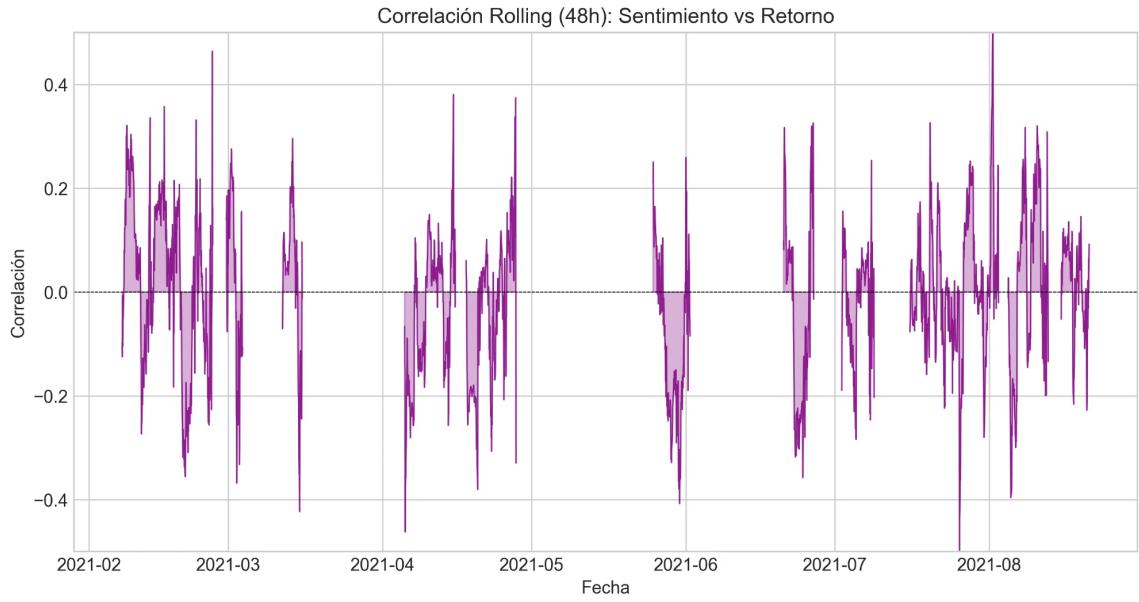


Figura 5: Correlación rolling (48 horas) entre sentimiento y retornos.

4.5. Cobertura social

El heatmap (Figura 6) muestra mayor actividad en horarios de mercado estadounidense.

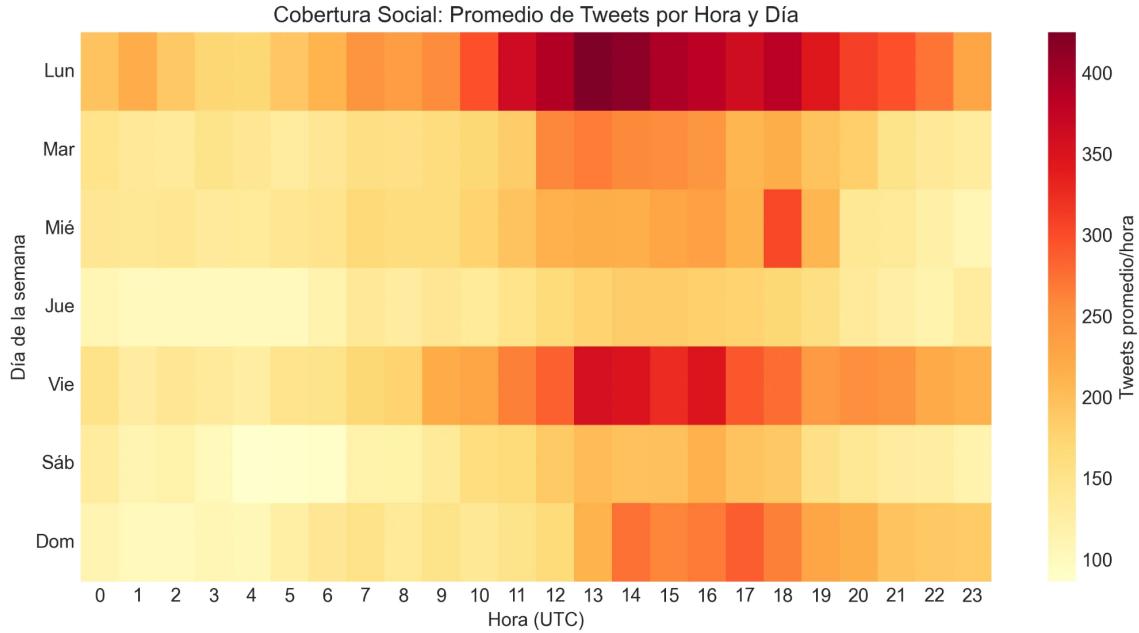


Figura 6: Cobertura social: promedio de tweets por hora y día.

4.6. Resultados de modelado

Cuadro 2: Resultados de regresión (Walk-Forward CV, 5 folds)

Horizonte	Modelo	Features	MAE	RMSE	R^2
1h	Naive Zero	—	0.0064	0.0089	-0.007
	Ridge	BASE	0.0066	0.0092	-0.056
	Random Forest	BASE	0.0080	0.0105	-0.383
	Random Forest	BASE+SENT	0.0073	0.0104	-0.644
6h	Naive Zero	—	0.0157	0.0217	-0.037
	Ridge	BASE	0.0204	0.0270	-0.644
	Random Forest	BASE	0.0221	0.0285	-0.927
	Random Forest	BASE+SENT	0.0191	0.0262	-0.642
24h	Naive Zero	—	0.0318	0.0405	-0.187
	Ridge	BASE	0.0575	0.0673	-2.594
	Random Forest	BASE	0.0569	0.0683	-2.901
	Random Forest	BASE+SENT	0.0509	0.0632	-2.105

4.7. Tests estadísticos

Cuadro 3: Resultados de tests estadísticos

Test	Estadístico	p-value
Wilcoxon (error BASE > error SENT)	13,028	0.683
Correlación Spearman (sent vs ret_24h)	$\rho = 0,088$	0.003**

** significativo al 1 %

4.8. Importancia de features

El análisis de permutation importance (Figura 7) muestra features de sentimiento entre las más relevantes individualmente.

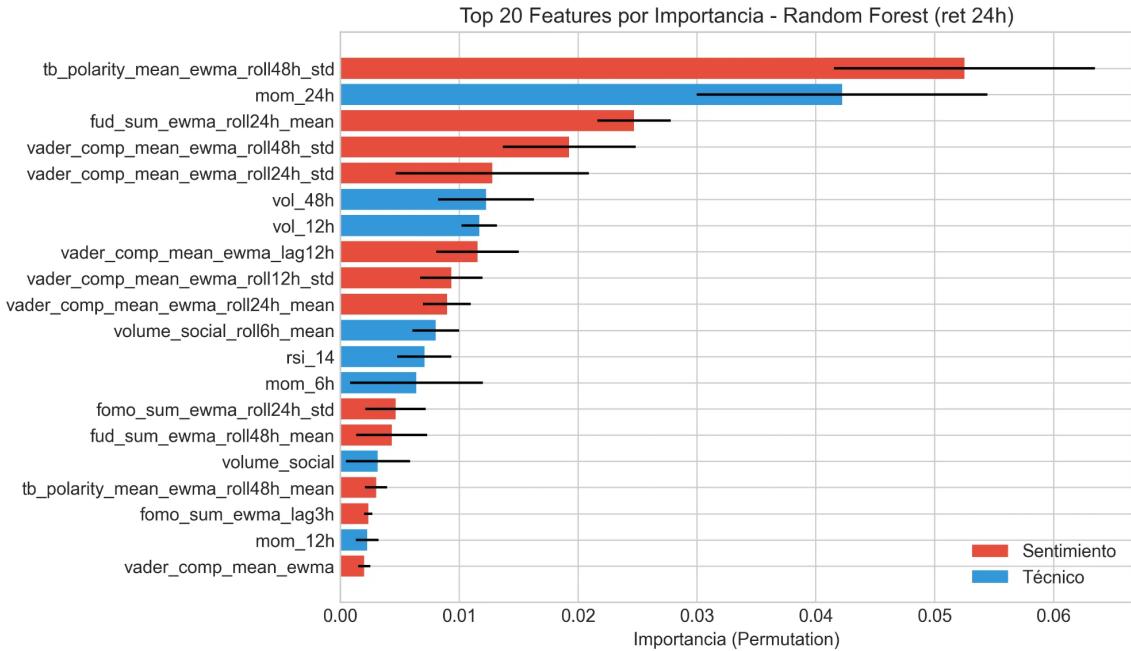


Figura 7: Top 20 features por importancia - Random Forest horizonte 24h.

Sin embargo, la importancia agregada (Figura 8) del grupo sentimiento es negativa (-0.048).

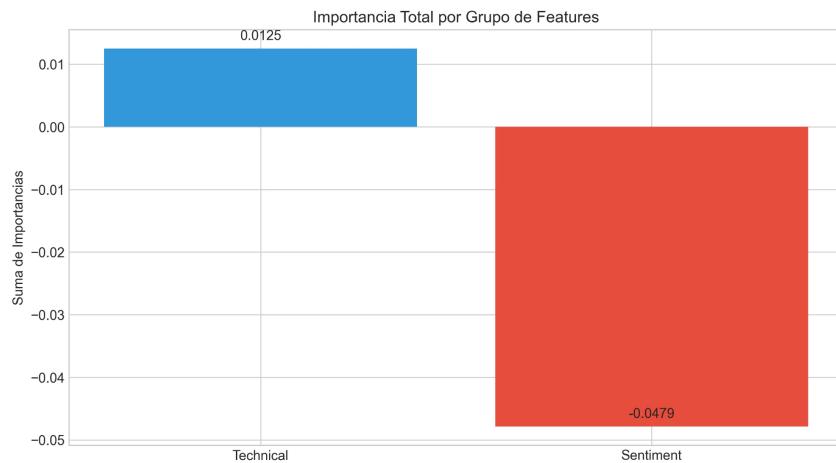


Figura 8: Importancia total por grupo de features.

4.9. Comparación BASE vs BASE+SENT

La distribución del cambio en error (Figura 9) está centrada ligeramente arriba de cero.

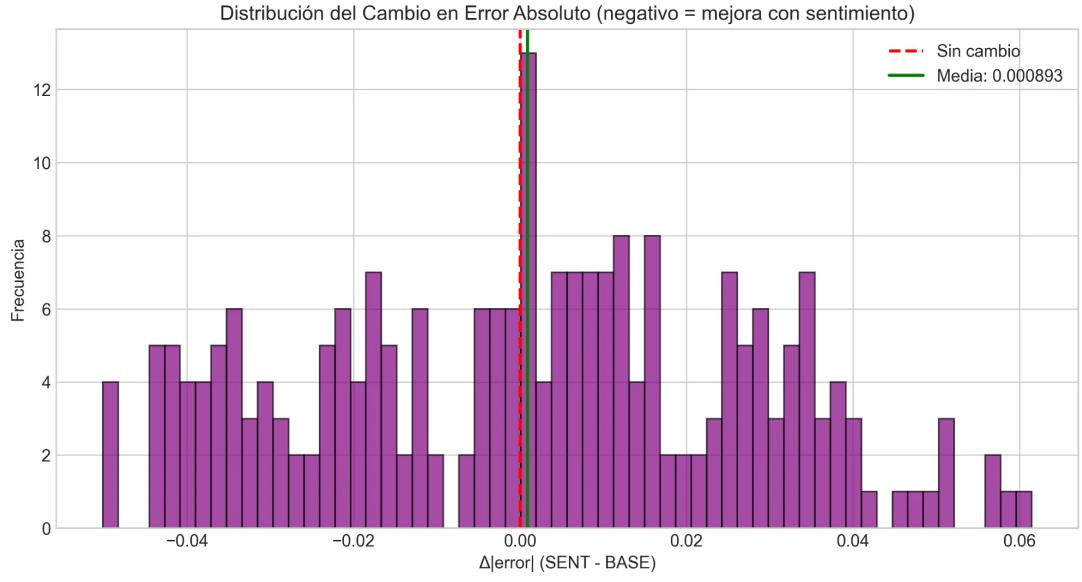


Figura 9: Distribución del cambio en error absoluto.

Los scatter plots (Figura 10) muestran R^2 negativo en ambos modelos.

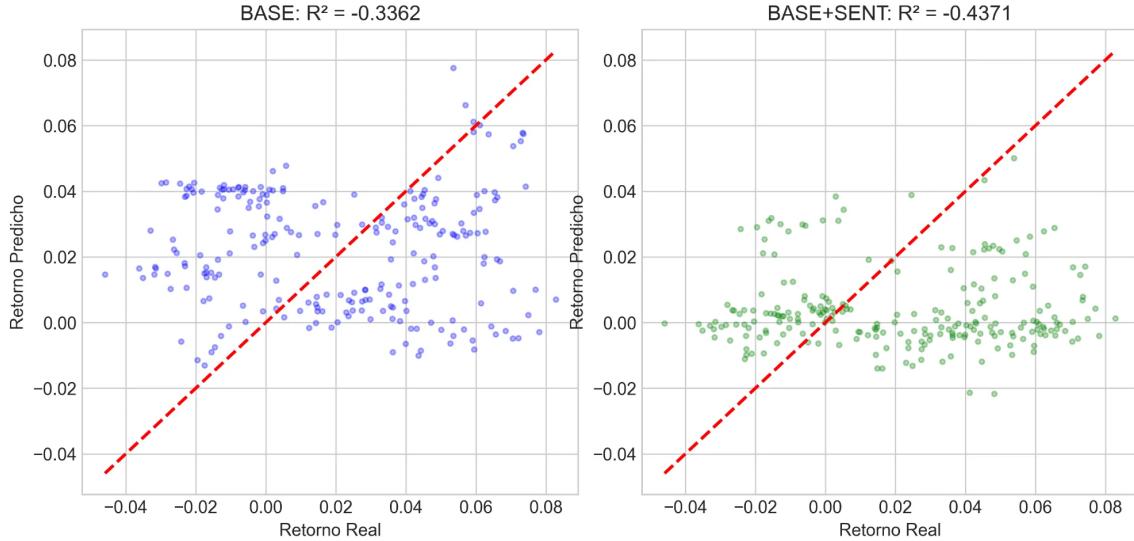


Figura 10: Predicciones vs retornos reales para modelos BASE y BASE+SENT.

4.10. Event study

El análisis de eventos (Figura 11) muestra tendencia alcista posterior a shocks de FUD.

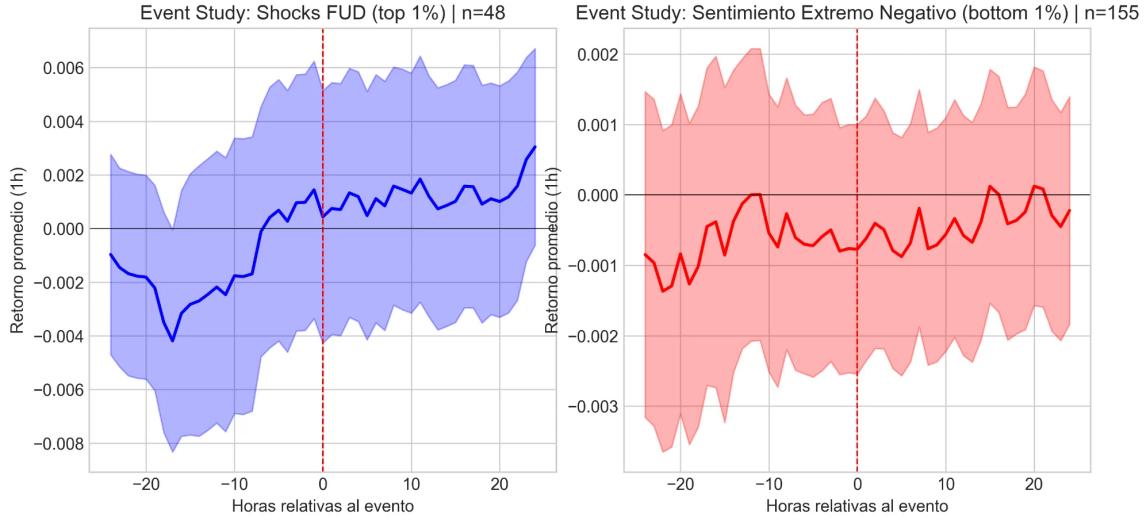


Figura 11: Event study: retornos alrededor de shocks de FUD y sentimiento negativo.

4.11. Modelo LSTM

El modelo LSTM (Figura 12) alcanzó MAE: 0.032 y R^2 : -0.451.

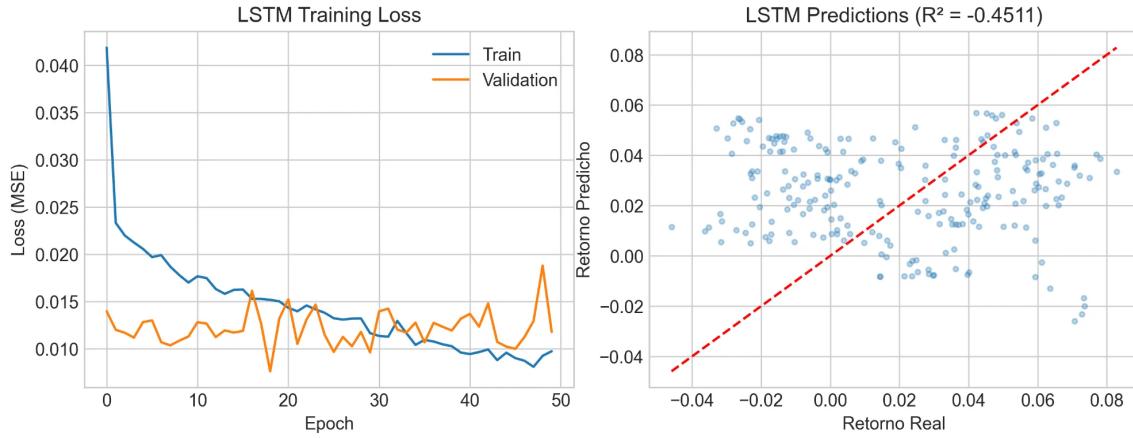


Figura 12: Resultados del modelo LSTM: curva de aprendizaje y predicciones.

5. Discusión

Los resultados no respaldan la hipótesis de que el sentimiento de Twitter mejora la predicción de precios de Bitcoin. A pesar de correlación significativa ($\rho = 0.088$, $p < 0.01$), esta relación es demasiado débil para traducirse en poder predictivo incremental.

Factores que explican estos hallazgos: (1) cobertura temporal limitada (32 %), (2) sesgo optimista estructural en Twitter crypto, (3) posible eficiencia del mercado, (4) ruido por bots y spam, (5) período con crash de mayo 2021.

5.1. Limitaciones

Período de 6 meses, uso exclusivo de Twitter, métodos léxicos limitados, ausencia de variables macroeconómicas.

6. Conclusiones

Este estudio evaluó la capacidad predictiva del sentimiento de Twitter sobre el precio de Bitcoin. Los principales hallazgos son:

1. Existe correlación significativa pero débil ($\rho = 0,088$, $p < 0,01$).
2. La incorporación de sentimiento no mejoró el desempeño (Wilcoxon $p = 0,68$).
3. El modelo naive supera a modelos complejos.
4. Las features de sentimiento tienen contribución agregada negativa.

Trabajo futuro: Incorporar FinBERT, expandir a Reddit, analizar períodos más extensos.

Referencias

1. Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia. *Scientific Reports*, 3, 3415.
2. Garcia, D., & Schweitzer, F. (2015). Social signals and algorithmic trading of Bitcoin. *Royal Society Open Science*, 2(9), 150288.
3. Kaminski, J. (2014). Nowcasting the Bitcoin market with Twitter signals. *arXiv:1406.7577*.
4. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in IR*, 2(1-2), 1-135.
5. Liu, B. (2012). Sentiment analysis and opinion mining. Morgan & Claypool.
6. Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis. *ICWSM 2014*.
7. Devlin, J., et al. (2019). BERT: Pre-training of deep bidirectional transformers. *NAAACL-HLT*.
8. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
9. Lamon, C., et al. (2017). Cryptocurrency price prediction using news and social media sentiment. *CS229 Stanford*.
10. Raheman, A., et al. (2022). Social media sentiment analysis for cryptocurrency market prediction. *arXiv preprint*.
11. Subramanian, H. (2024). A decision support system using social media signals with LSTM. *Decision Support Systems*.
12. Bouteska, A. (2022). Predictive power of investor sentiment for Bitcoin returns. *The Social Science Journal*.