
EFFECTO DEL SENTIMIENTO EN REDES SOCIALES SOBRE EL PRECIO DEL BITCOIN

Sebastián Marinovic Leiva
Ingeniero metalúrgico y civil industrial
Alumno Magíster Data Science
Universidad de Las Américas
sebamarinovic.leiva@gmail.com

Ricardo Iván Lizana Aseña
Ingeniero en informática
Alumno Magíster Data Science
Universidad de Las Américas
ricardo.lizana02@gmail.com

Luis Andrés Gutiérrez González
Ingeniero civil industrial
Alumno Magíster Data Science
Universidad de Las Américas
luis.gutierrezgo@mail.udp.cl

December 18, 2025

ABSTRACT

Bitcoin es un activo altamente volátil cuyo comportamiento puede verse influenciado por factores exógenos, entre ellos el sentimiento expresado por los inversionistas en redes sociales. Investigaciones previas han mostrado que señales sociales extraídas de plataformas digitales pueden complementar variables financieras tradicionales, aportando información útil para predicción [1, 2]. Este proyecto analiza si el sentimiento publicado en Twitter posee capacidad predictiva sobre variaciones del precio de Bitcoin en una ventana de 24 horas. Para ello, se emplean técnicas de procesamiento de lenguaje natural (NLP) mediante enfoques léxicos y modelos basados en Transformers para cuantificar el sentimiento [3, 4], integrándolo luego en modelos predictivos tradicionales y redes neuronales recurrentes LSTM [5]. Se espera que la incorporación de variables de sentimiento mejore la precisión frente a modelos basados únicamente en datos históricos de precios.

1 Introducción

Bitcoin (BTC) es el activo digital más representativo del mercado de criptomonedas, caracterizado por alta volatilidad y sensibilidad a factores exógenos. A diferencia de los mercados tradicionales, su dinámica puede estar condicionada por señales de atención pública y percepción colectiva [6]. Con el auge de redes sociales, los mercados reaccionan en tiempo real a información difundida en plataformas como Twitter y Reddit, y publicaciones virales pueden coincidir con movimientos de precio, especialmente en activos altamente especulativos.

La literatura ha explorado el uso de Twitter como fuente de señales: se han reportado asociaciones entre indicadores sociales y dinámica del mercado, y estrategias de trading algorítmico basadas en señales sociales [2, 1]. En paralelo, el análisis de sentimiento es una línea consolidada en NLP para cuantificar el tono emocional de textos [7, 8]. Este trabajo integra ambas áreas para evaluar la capacidad predictiva del sentimiento sobre el precio de Bitcoin.

2 Descripción de tareas y construcción de datos

Los datos utilizados provienen de dos fuentes principales:

- **Tweets:** conjunto de datos público con publicaciones relacionadas con Bitcoin (por *keywords*), con marca temporal asociada a cada tweet.

- **Precio BTC:** serie OHLC (*open, high, low, close*) obtenida desde una fuente de mercado para el mismo período, asegurando consistencia temporal.

Posteriormente se ejecuta limpieza y preprocesamiento, incluyendo normalización de texto, manejo de valores faltantes y sincronización temporal. Finalmente, las variables se agregan a una frecuencia uniforme (por ejemplo, diaria u horaria) para construir un dataset multivariable apto para modelado de series temporales.

3 Planteamiento del problema

A pesar del creciente interés en el uso de redes sociales como fuente de información financiera, la evidencia empírica sobre su real poder predictivo sigue siendo heterogénea. Mientras algunos trabajos reportan mejoras al incorporar sentimiento social en modelos de predicción [9, 10], otros señalan que dichas mejoras dependen del período analizado y del modelo aplicado [11].

La pregunta central de esta investigación es:

¿Puede el sentimiento expresado en redes sociales como Twitter anticipar movimientos significativos en el precio futuro del Bitcoin, y esta señal mejorar la exactitud de modelos predictivos de series temporales?

4 Objetivos

4.1 Objetivo general

Evaluar si cambios en el sentimiento promedio extraído de Twitter pueden predecir variaciones en el precio de Bitcoin en ventanas futuras (por ejemplo, 24 horas) y si dicha señal mejora la precisión de modelos predictivos tradicionales y avanzados.

4.2 Objetivos específicos

- Construir métricas cuantitativas de sentimiento desde texto social usando enfoques léxicos y modelos basados en Transformers [3, 4].
- Analizar la distribución temporal del sentimiento y su relación con series de precio.
- Comparar modelos con y sin variables de sentimiento:
 1. Modelos tradicionales: regresión lineal, Random Forest.
 2. Modelos neuronales: LSTM integrando sentimiento como *feature* [5].
- Evaluar desempeño con métricas MAE, MSE y R^2 .
- Visualizar resultados para interpretar la relación sentimiento–precio.

5 Marco Teórico

5.1 Bitcoin y volatilidad

La volatilidad de BTC se asocia a factores macro, condiciones de liquidez y señales de atención pública. Se ha documentado relación entre atención (Google Trends/Wikipedia) y dinámica de BTC [6]. La Figura 1 muestra la evolución del precio de cierre diario y evidencia variaciones abruptas y cambios de tendencia en períodos cortos, motivando el uso de variables exógenas como el sentimiento social.

La Figura 1 evidencia un comportamiento altamente variable, coherente con lo reportado en la literatura [6]. Este patrón refuerza la hipótesis de que factores no tradicionales, como la percepción colectiva y el sentimiento expresado en redes sociales, pueden desempeñar un rol relevante en la dinámica del precio, especialmente en horizontes temporales cortos.

5.2 Análisis de sentimiento en finanzas

El análisis de sentimiento busca cuantificar el tono emocional de un texto (positivo/negativo/neutro) y ha sido ampliamente estudiado en NLP [7, 8]. En redes sociales, donde los textos son breves e informales, modelos como VADER han mostrado buen desempeño en la estimación de polaridad [3]. En enfoques modernos, Transformers como

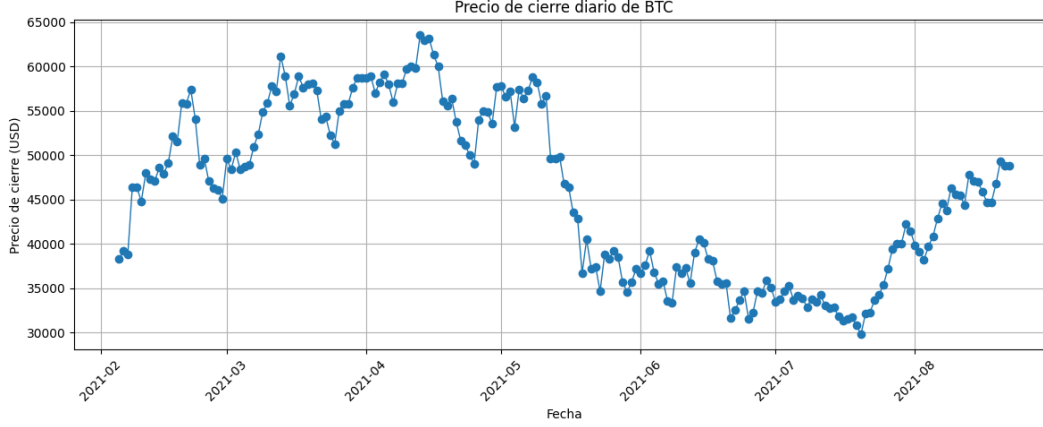


Figure 1: Precio de cierre diario de Bitcoin (BTC) durante el período analizado.

BERT capturan contexto bidireccional y suelen mejorar el desempeño en tareas de NLP [4]. La Figura 2 presenta la distribución del sentimiento en tweets analizados.

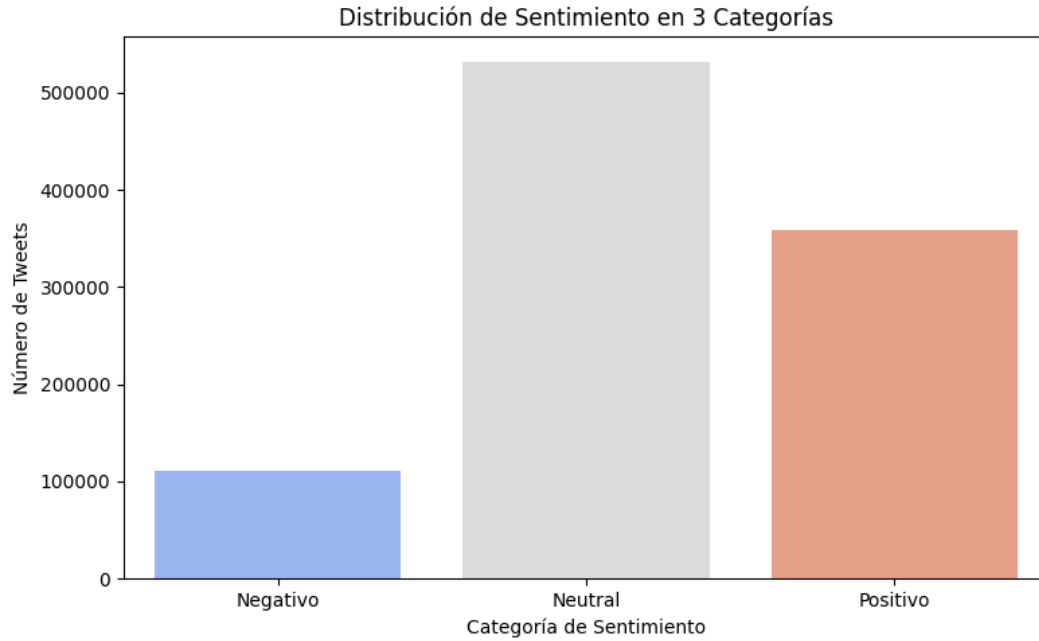


Figure 2: Distribución del sentimiento estimado en tweets relacionados con Bitcoin.

En la Figura 2 se observa predominio de sentimiento neutral, consistente con hallazgos previos donde una parte relevante del contenido informativo no expresa polaridad fuerte [7, 8]. Sin embargo, la presencia de componentes positivos y negativos sugiere períodos de mayor carga emocional que podrían asociarse a eventos de mercado y aportar señal adicional.

5.3 Relación preliminar entre sentimiento y precio

El análisis exploratorio sugiere que variaciones extremas en el sentimiento promedio tienden a concentrarse en períodos de alta volatilidad del precio. Este patrón ha sido reportado en investigaciones previas, donde el sentimiento social actúa como una señal complementaria más que como un predictor aislado [1, 2]. En consecuencia, este trabajo considera el sentimiento como una variable adicional dentro de un modelo multivariado.

5.4 Series temporales y LSTM

Las redes neuronales recurrentes, y en particular las arquitecturas LSTM, han demostrado ser efectivas para modelar dependencias de largo plazo en series temporales [5]. Estudios recientes han aplicado LSTM a la predicción de precios de criptomonedas incorporando variables de sentimiento social [12, 13].

6 Hipótesis

H1: El sentimiento social promedio positivo en Twitter se correlaciona con incrementos en el precio de Bitcoin en las siguientes 24 horas.

H0: No existe relación estadísticamente significativa entre el sentimiento social y la variación posterior del precio de Bitcoin.

7 Metodología implementada

7.1 Recolección y preprocesamiento

- Tweets: dataset con columnas típicas *date*, *text* (y/o *cleaned_text*), *sentiment*.
- Precio BTC: serie OHLC con *date*, *open*, *high*, *low*, *close*, *volume*.
- Limpieza de texto: eliminación de URLs, menciones, hashtags; normalización; tokenización.
- Sincronización: agregación temporal (p. ej. diaria u horaria) para alinear sentimiento y precio.

7.2 Análisis de sentimiento

Se calcula una métrica de sentimiento por texto mediante un baseline léxico (por ejemplo VADER) [3] y se contempla una extensión con Transformers (BERT/FinBERT) para mejorar la precisión contextual [4]. Posteriormente, el sentimiento se agrega temporalmente (promedio y/o cuantiles por ventana) para integrarse como *feature*.

7.3 Modelos de predicción

7.3.1 Modelos tradicionales

- Regresión lineal.
- Random Forest Regressor (baseline no lineal).

7.3.2 Modelo neuronal recurrente

Se implementa una red LSTM con regularización (Dropout) para capturar dependencias temporales [5]. La entrada considera secuencias multivariantes (precio pasado + sentimiento agregado) y la salida corresponde al precio futuro (horizonte 24 horas o el paso temporal definido).

7.3.3 Preprocesamiento para modelado

- Escalado: MinMaxScaler.
- Ventanas temporales: secuencias para LSTM.
- Partición: train/test (y opcionalmente validación) respetando el orden temporal.

7.3.4 Métricas de evaluación

MAE, MSE y R^2 para comparación consistente entre modelos.

8 Resultados observados (análisis exploratorio)

En el análisis exploratorio se observa que:

- La polaridad promedio tiende a concentrarse cerca de 0 con componentes positivos y negativos.
- Existen picos de sentimiento extremo potencialmente asociados a eventos del mercado.

- En modelado preliminar, LSTM tiende a capturar mejor dependencias temporales que modelos tradicionales, y el sentimiento puede aportar señal adicional en ciertos períodos [12, 13].

9 Limitaciones identificadas

- Sesgo temporal del período analizado.
- Ruido textual por bots/spam y contenido no relevante.
- Limitaciones de métodos léxicos (sarcasmo, jerga del dominio) [8].
- Sincronización temporal y rezagos de reacción entre redes y mercado.

Trabajo futuro

- Incorporar Transformers específicos de finanzas (FinBERT) y comparar contra baselines.
- Expandir a Reddit u otras plataformas.
- Probar ventanas predictivas adicionales y variables exógenas.

References

- [1] David Garcia and Frank Schweitzer. Social signals and algorithmic trading of bitcoin. *Royal Society Open Science*, 2(9):150288, 2015.
- [2] Jacek Kaminski. Nowcasting the bitcoin market with twitter signals. *arXiv preprint arXiv:1406.7577*, 2014.
- [3] Clayton J Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *ICWSM*, 2014.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT*, 2019.
- [5] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [6] Ladislav Kristoufek. Bitcoin meets google trends and wikipedia. *Scientific Reports*, 3:3415, 2013.
- [7] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135, 2008.
- [8] Bing Liu. *Sentiment analysis and opinion mining*. Morgan & Claypool, 2012.
- [9] Christiaan Lamon, Erik Nielsen, and Esteban Redondo. Cryptocurrency price prediction using news and social media sentiment. *CS229 Machine Learning*, 2017.
- [10] A Raheman, A Kolonin, I Fridkins, and I Ansari. Social media sentiment analysis for cryptocurrency market prediction. *arXiv preprint*, 2022.
- [11] Ahmed Bouteska. Predictive power of investor sentiment for bitcoin returns. *The Social Science Journal*, 2022.
- [12] H Subramanian. A decision support system using signals from social media with an lstm network to predict bitcoin prices. *Decision Support Systems*, 2024.
- [13] D Grubišić, B Škrabić-Perić, and M Jadrić. Bitcoin price prediction using sentiment analysis. *ST-OPEN*, 2025.