ST open

RESEARCH ARTICLE

Bitcoin price prediction using sentiment analysis

Dora Grubišić[1,2], Blanka Škrabić Perić[3], Mario Jadrić[4]

[1]Faculty of Economics, University of Split, Split, Croatia

[2]OTP banka d.d., Split, Croatia

[3]Faculty of Economics, Department of Quantitative Methods, University of Split, Split, Croatia

[4]Faculty of Economics, Department of Business Informatics, University of Split, Split, Croatia

Correspondence to:
Mario Jadrić
University of Split Faculty of Economics,
Department of Business Informatics
Cvite Fiskovića 5, 21000 Split, Croatia
jadric@efst.hr

Cite as:
Grubišić D, Škrabić-Perić B, Jadrić M. Bitcoin price prediction using sentiment analysis. ST-OPEN. 2025;6:e2025. 2301.9.

DOI:
https://doi.org/10.48188/so.6.12

Aim: To explore the causal relationship between social media sentiment, related behavioral factors, and bitcoin price performance, and to develop predictive models with higher accuracy in forecasting the bitcoin price by incorporating sentiment analysis.

Methods: We used the Valence Aware Dictionary and sEntiment Reasoner module to perform a sentiment analysis of 896,464 Twitter posts (tweets) published between November and December 2021, which we collected *via* web scraping. We created several forecasting models using the average daily sentiment polarity, the average daily number of tweets, and search interest for “bitcoin” on Google and Wikipedia as input variables. We predicted future bitcoin prices using vector autoregression (VAR), Prophet, and long short-term memory (LSTM) artificial neural network models and evaluated their predictive accuracy using the mean absolute percentage error (MAPE) as a performance measure.

Results: The results suggest a Granger causal relationship between social media sentiment and bitcoin prices. The standard VAR model achieved a MAPE of 8%, while the LSTM model had a lower error rate of 5%. The Prophet model had a MAPE of 11%.

Conclusion: Our results underline the highly speculative nature of bitcoin, especially in times of high prices. The inclusion of behavioral variables in the development of bitcoin price prediction models significantly improved their prediction accuracy, with the LSTM neural network model proving to be an extremely effective tool in this sense.

Keywords: behavioral factors; bitcoin; LSTM artificial neural network model: Prophet; sentiment analysis; vector autoregression

Introduction

Launched in 2008, bitcoin was conceived as a secure digital payment system that enables peer-to-peer transactions without traditional intermediaries (*1*). Although conceived as a virtual currency, its exact definition remains unclear, while its limited adoption means that transaction volumes are low compared to traditional methods and that there are no safeguards or backing from banks or governments (*2*-*4*). Consequently, it does not fulfill the core functions of money and cannot be considered a real currency. However, Vora (*5*) predicts that wider adoption could strengthen bitcoin's role as a unit of account due to its convertibility into national currencies. In contrast, Kubát (*6*) rejects bitcoin as money from a theoretical, empirical, and legal point of view. Despite the lack of cash flows or real utility (*7*), bitcoin's volatility and potential returns attract speculative investors (*8*, *9*). Such investor behavior is at odds with conventional financial markets, where price fluctuations and instability are typically undesirable (*10*).

Since technical and fundamental factors cannot fully explain bitcoin's strong price fluctuations (*11*, *12*), here we examine how alternative, behavioral factors influence its price formation. Existing literature highlights the impact of various economic, financial, and behavioral factors on bitcoin's valuation. Poyser (*12*, *13*) found that the exchange rates of the US dollar and the euro positively affect bitcoin's price, as does strong investor interest. Ciaian and colleagues (*14*) observed that demand, measured by transaction volume, is a key factor in mature cryptocurrency markets, while investor interest and new information dominated in the early stages of bitcoin. Zhu and colleagues (*15*) showed that macroeconomic and stock market variables such as the consumer price index, the US dollar index, the Dow Jones index, Federal Reserve System interest rates, and the price of gold have a significant impact on bitcoin, with the link to the US dollar being the strongest.

Behavioral factors significantly impact the price movements of virtual currencies, prompting researchers to integrate them into prediction models. In contrast to traditional financial markets, cryptocurrency markets are dominated by retail investors who often lack expertise and rely heavily on news, social media, and public opinion when making investment decisions (*16*, *17*). As its user base includes the wider public, as well as governments, banks, funds, and privately-owned businesses, user-generated content on platforms such as Twitter provides valuable insights into market sentiment, which can be quantified using text mining and sentiment analysis to identify emotions and determine polarity (*18*). This behavioral data complements traditional market indicators and provides a deeper understanding of the psychological forces driving price fluctuations. In recent years, sentiment analysis has become increasingly important in financial forecasting, especially in predicting the price performance of virtual currencies (*19*-*29*). In many studies, bitcoin price prediction is considered as a classification problem, which is why they used machine learning and deep learning algorithms to achieve relatively high accuracy (*23*–*29*). However, the extent to which sentiment directly influences bitcoin's price is difficult to generalize, as the results are inconsistent across different time periods. Here we focus on bitcoin's all-time high and early decline in November and December 2021, a period characterized by increased speculation and volatility.

We use sentiment analysis with statistical and machine learning models to examine whether public interest and polarity of tweets influence price movements, whether behavioral factors can improve prediction accuracy, and whether machine learning models outperform traditional statistical approaches. While existing literature often examined the relationships between cryptocurrencies and traditional currencies using generalized autoregressive conditional heteroskedasticity models or machine learning (*23*–*31*), other studies focused on cultural and psychological factors using surveys and structural equation modeling (*8*, *32*). This study adds value by creating a sentiment index to measure the influence of media on bitcoin prices and combining machine learning with vector autoregression (VAR), which, unlike the autoregressive integrated moving average, captures dynamic interdependencies between time series and is therefore ideal for analyzing complex financial relationships (*33*).

Methods

We based our bitcoin price prediction on two behavioral factors: indicators of public opinion and investment attractiveness. Besides a traditional VAR model, we used an automated regression model (Prophet) and a long short-term memory (LSTM) deep learning model in the analysis. We conducted the sentiment analysis using an original dataset collected during periods of significant upward and downward price fluctuations of bitcoin, rather than data from secondary sources.

Data collection and preparation

Based on favorable literature outcomes (*23*–*25*, *27*), we sourced sentiment analysis data from Twitter. Specifically, we collected Twitter posts using “snscrape”, a command-line web scraping tool chosen for its ease of use, access to historical data, and fewer limitations compared to the Twitter API (*31*). Snscrape does not require a Twitter account and allows flexible extraction of data points such as hashtags, tweets, users, communities, and cashtags. Without limiting the number of posts, we collected 1.64 million tweets with the hashtag “bitcoin” published between November and December 2021, a time when bitcoin prices peaked. After cleaning and pre-processing with Python’s “pandas” library, 896,464 observations remained for analysis.

For the purposes of sentiment analysis, textual data needs to be processed prior to machine interpretation and analysis in a different manner than in conventional data mining, as semantic factors such as polysemy, grammar and spelling errors, and informal language can significantly affect its results. Here we used Python, version 3.10.7 (Python Software Foundation, Wilmington, Delaware, USA) for data cleaning, transformation, analysis, and model development (*34*). We performed the sentiment analysis using the Valence Aware Dictionary and sEntiment Reasoner (VADER) module because of its ability to handle and interpret informal language, emoticons, and punctuation in social media easily, quickly, and accurately (*35*).

We began data cleaning by removing zeros, missing and duplicate observations, leaving only the tweet’s text and date intact. We excluded non-English posts due to VADER’s lack of multi-language support. Text preprocessing was handled through the “NLTK” and “regex” librarires, including the removal of stop words (except negations such as “not” and “very”), hyperlinks, hashtags, symbols, and numbers. The tweets were tokenized for machine processing. Normalization of upper and lower case was avoided, as VADER is case-sensitive to improving polarity detection, which reflects human perception of text emphasis (*35*).

Sentiment analysis

The VADER module is a simple and effective analysis tool that only requires the preconfigured SentimentAnalyzer object and the desired text data as input. The methodological framework for sentiment analysis used in this study is shown in **Figure 1**. Following the sentiment analysis, VADER creates an overall score with percentages for positive, negative, or neutral sentiments in the text, as well as a composite polarity score. Sentiment is computed based on composite (aggregate) polarity values, with the following thresholds:

• Negative sentiment for values < -0.5,

• Neutral sentiment for values between -0.5 and 0.5,

• Positive sentiment for values > 0.5.

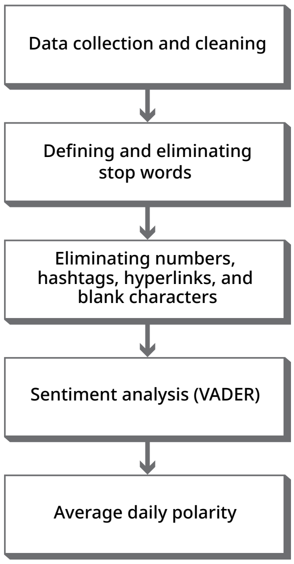

Figure 1. Overview of the sentiment analysis methodology and workflow. The process involves data collection and preprocessing in accordance with text analysis principles, sentiment analysis, and calculation of daily average sentiment polarity scores to generate a time-series representation.

We calculated the average daily sentiment using the average daily compound polarity score. As previous studies show a strong correlation between tweet volume and bitcoin prices (*23*), daily tweet volume was also calculated and included in the model as a measure of public interest. Like Abraham and colleagues (*23*), we included Google Trends data on bitcoin search, which show a stronger correlation with price fluctuations than sentiment alone (*36*), Wikipedia search activity for “bitcoin”, obtained *via* Pageviews Analytics (*37*), and Yahoo Finance historical bitcoin prices, with the closing prices of the day as the target variable (*38*).

Model selection

The final dataset, structured as a time series, required appropriate forecasting methods for price prediction. We used a multivariate VAR model, the Prophet machine learning-based forecasting tool, and an LSTM model for this purpose, as they are commonly used in similar studies (*24*, *25*, *29*, *39*, *40*) (**Figure 2**).

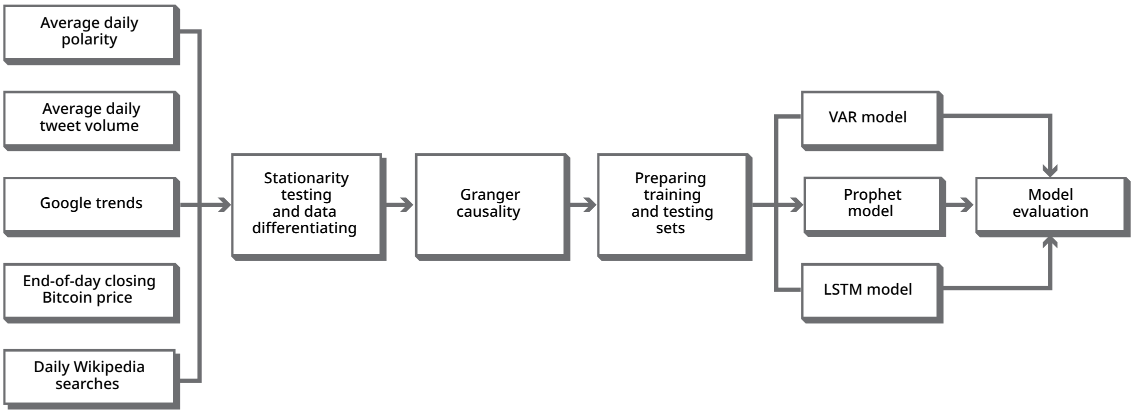

Figure 2. Illustration of the bitcoin price prediction workflow. The data collected from various sources was tested for stationarity and differentiated if necessary. After testing for Granger causality, the data was further processed to construct VAR, Prophet and LSTM models.

VAR models are widely used in financial analyses to examine and forecast multivariate time series by modeling the variables and their interrelationships as systems of equations (*33*). Each variable is expressed as a linear function of its own past lags and those of the past lags of other variables. Stationarity is required, and the optimal lag length is determined using criteria such as the Akaike information criterion, the Bayesian information criterion, or the Hannan-Quin information criterion. A precise selection of lags is crucial, as too many lags increase prediction error, while too few cause autocorrelation (*41*). Since stationarity of time series is a prerequisite for estimating VAR models and performing Granger causality tests, the variables under consideration must be tested to determine whether this condition is met. Although stationarity is not essential for other modeling approaches used in this analysis, it still significantly affects the prediction accuracy. We tested stationarity using the augmented Dickey-Fuller (ADF) unit root test (**Table 1**).

Table 1. ADF test results for variables in the bitcoin price prediction dataset*

Variable	ADF	*P*-value	Lag	n	1%	5%	10%
Btc_price	-0.89	0.792	0	60	-3.54	-2.91	-2.59
Sentiment	-1.54	0.523	1	59	-3.55	-2.91	-2.59
Tweet_volume	-1.79	0.385	1	59	-3.55	-2.91	-2.59
Google_trends	-4.66	0.000	0	60	-3.55	-2.91	-2.59
Wikipedia	-2.05	0.264	0	60	-3.54	-2.91	-2.59

*Abbreviations: ADF – augmented Dickey-Fuller test, btc_price – daily closing price of bitcoin.

VAR models offer higher prediction accuracy and better identification of relationships between variables (*33*), making them suitable for predicting bitcoin prices. After the VAR evaluation, we used the Granger causality test to examine the influence of sentiment and public interest. The Python libraries "statsmodels" and "scikit-learn" were used for the evaluation.

Prophet is a specialized tool for time series forecasting that automates predictions and handles data preprocessing, outliers, and missing values (*42*). It incorporates vacation data to recognize seasonality and trends, and users can add custom holidays to account for exogenous shocks. The Prophet's ability to detect turning points or abrupt trend shifts improves forecast accuracy by estimating spurious events, while the automatic model selection and configuration save time and make it accessible to non-experts.

Recurrent neural networks are designed for sequential data such as time series but have difficulty maintaining long-term dependencies (*43*). LSTM networks overcome this limitation and improve prediction accuracy. We developed our LSTM model for time series prediction using the "TensorFlow" library with the "Keras" interface.

Results

Based on the obtained *P*-values, we found that the bitcoin price, tweet sentiment, tweet volume, and Wikipedia activity were not stationary in their levels. In contrast, the Google Trends variable exhibited stationarity, with a corresponding *P*-value <0.001. To address this non-stationarity, we transformed all variables by first-order difference, after which we reapplied the ADF test and found that, according to the obtained *P*-values (**Table 2**), all variables became stationary, making further testing unnecessary. We then used this first-differenced data as input for subsequent model development and prediction.

Table 2. ADF test results for first-differenced variables in the bitcoin price prediction dataset*

Variable	ADF	*P*-value	Lag	n	1%	5%	10%
Btc_price	-6.82	<0.001	1	58	-3.55	-2.91	-2.59
Sentiment	-7.83	<0.001	1	58	-3.55	-2.91	-2.59
Tweet_volume	-10.34	<0.001	0	59	-3.55	-2.91	-2.59
Google_trends	-10.09	<0.001	0	59	-3.55	-2.91	-2.59
Wikipedia	-6.07	<0.001	3	56	-3.55	-2.91	-2.59

*Abbreviations: ADF – augmented Dickey-Fuller test, btc_price – daily closing price of bitcoin.

VAR model

To evaluate the prediction accuracy, we divided the dataset into a training dataset and a test dataset with a prediction horizon of 10 days. By differentiating the data, the sample size was reduced by one, resulting in 50 training observations. The VAR model, estimated using the "statsmodels" library, used the first differenced data. Statsmodels automatically selects the optimal lag length, requiring a maximum lag limit and an information criterion. Using the Akaike information criterion, which is recommended for small samples, the optimal lag was set to six. The model was evaluated using the VAR function. Prior to prediction, diagnostic tests were used to assess the autocorrelation of the residuals, normality, and stability of the model. The Durbin-Watson test was performed to determine the autocorrelation in the residuals. The results are summarized in **Table 3**.

Table 3. Results of the Durbin-Watson test for autocorrelation in the residuals*

Variable	Durbin-Watson test (dL = 1.164, dU = 1.587)
Btc_price	1.85
Sentiment	2.15
Tweet_volume	1.96
Google_trends	1.99
Wikipedia	1.85

*Abbreviations: btc_price – daily closing price of bitcoin, dL – lower test bound, dU – upper test bound.

Since all values lie within the reference interval from the statistical tables for the critical Durbin-Watson values, it can be concluded that the estimated model does not exhibit any significant residual autocorrelation. The multivariate normality of the residuals was assessed using the Jarque-Bera test. The test results were below the critical value, indicating that the residuals were normally distributed. The stability of the model was checked using a stability test from the statsmodels library, which showed that all eigenvalues of the companion matrix in the VAR (*6*) model are less than one in absolute value. After estimating the VAR model, the Granger causality test was performed to determine a possible correlation between the past values of the behavioral factors and the bitcoin price movements.

Based on the *P*-values of the Granger causality test (only the lowest are shown in **Table 4** for simplicity), the sentiment variable appears to have a Granger causality for the bitcoin price. No causality was found for the remaining variables, which serve as measures of investor interest or investment attractiveness. Finally, the estimated VAR model was used to generate a 10-day forecast for the bitcoin price (**Figure 3**). The predicted values were then transformed back to their original values for comparison with the actual values and to evaluate the performance of the model.

Table 4. Results of the Granger causality test between variables and bitcoin prices*

Variables (x, y)	P-value
Sentiment→btc_price	0.027
Tweet_volume→btc_price	0.350
Google_trends→btc_price	0.602
Wikipedia→btc_price	0.076

*Abbreviations: btc_price – daily closing price of bitcoin.

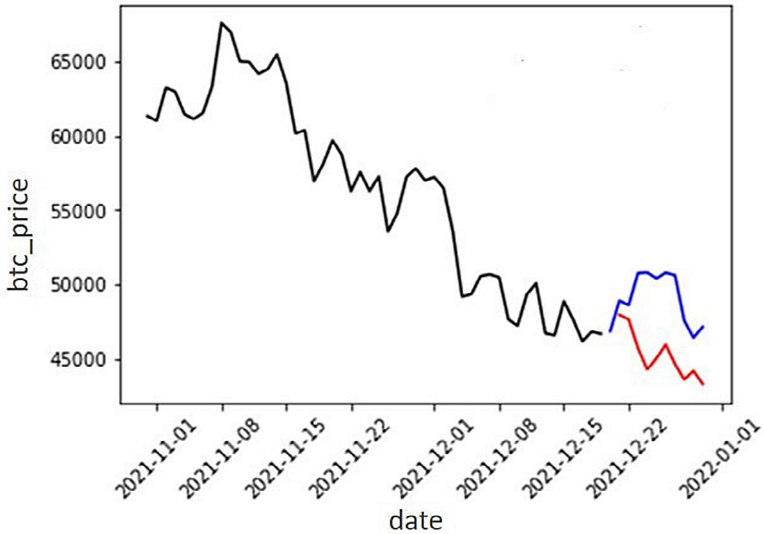

Figure 3. Bitcoin (Btc) price forecast based on the VAR model. The black line represents training data, the blue line indicates real-world prices, and the red line shows predicted values.

Prophet

Although the Prophet supports parameter customization, we made no changes when evaluating the accuracy of the automated predictions for comparison with other models. We used pre-differentiated data without further transformations for the prediction of bitcoin prices, as the Prophet automatically detects and corrects irregularities. We integrated all relevant variables into the model, which we then trained on the intended data set. The price forecasts were created with the calibrated model and compared with real prices with inverse differentiating (**Figure 4**). The results showed a significant discrepancy over the ten-day forecast horizon, with Prophet consistently underestimating bitcoin prices compared to actual market values.

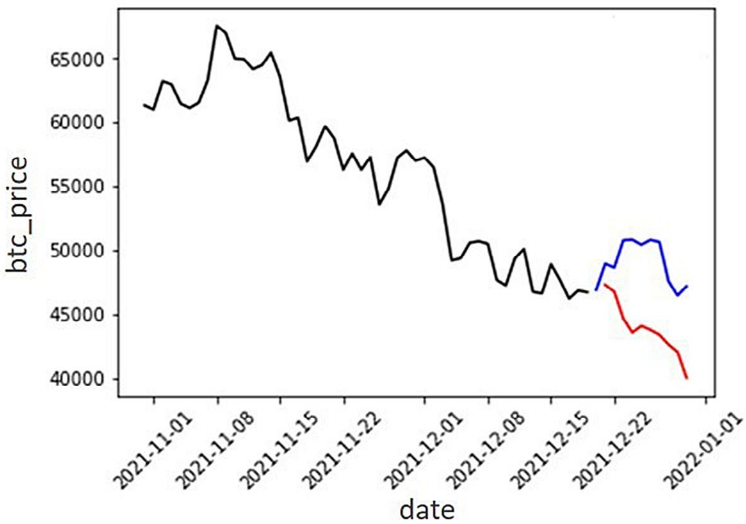

Figure 4. Bitcoin (Btc) price prediction results using the Prophet model. The black line represents training data; the blue line represents real-world prices, and the red line indicates predicted prices.

LSTM neural network model

Before creating the LSTM model, we standardized the initially differenced data using "scikit-learn" to ensure consistent scaling of variables, which is crucial for multivariate machine learning models. The model was developed by trial and error and tuning the hyperparameters to balance simplicity and predictive accuracy. Starting from a base architecture, the complexity gradually increased while paying attention to overfitting and underfitting. To mitigate these risks, we added a dropout layer with a probability of 0.2, randomly deactivating 20% of the neurons during the test. Additionally, 15% of the data was set aside as a validation set to track the model's performance and identify errors caused by excessive depth or complexity. Adjustments were made to achieve a simpler, more efficient structure. After testing different neuron counts and layer configurations, the final LSTM network architecture included three layers: an LSTM layer, a dropout layer, and a dense layer.

We evaluated the LSTM model using the mean absolute percentage error (MAPE), which we selected for its interpretability, its compatibility with positive values, and its ability to compare model deviations (*44*). We monitored the training and validation losses to control overfitting and underfitting, with the validation loss shown in **Figure 5**.

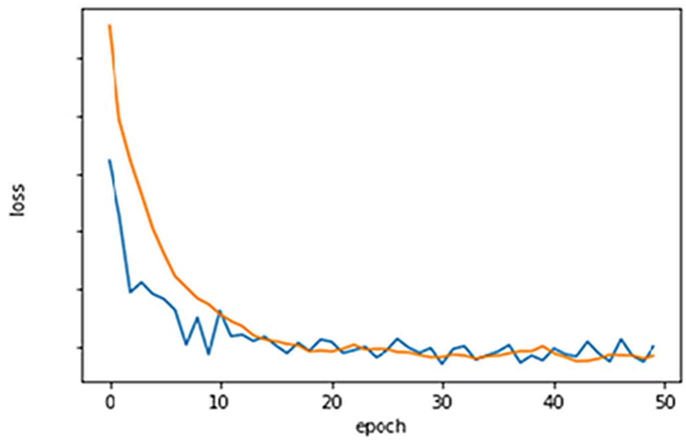

Figure 5. Loss functions. The training loss function is shown in blue, while the validation loss function is shown in orange.

Beyond model evaluation, we used the relationship between training and validation loss functions to determine the optimal number of epochs for the model, with loss minimization observed at 50 epochs. Before evaluating the prediction accuracy, we performed inverse standardization to compare the predictions with real data. For the evaluation, the data was reset to its original scale. predicted prices deviate slightly from the actual prices, but the difference is smaller in comparison to the VAR and Prophet models (**Figure 6**). It is noticeable that the overfitting was reduced to a maximum compared to alternative models, as the predicted values correspond better to the real values.

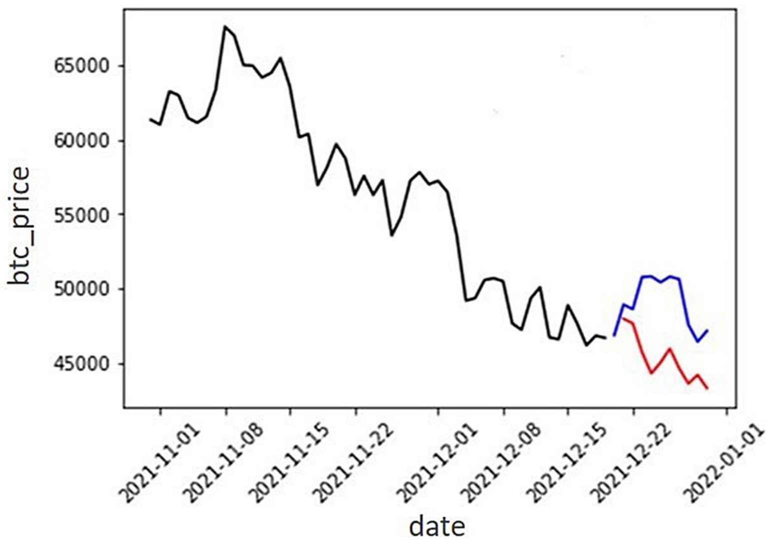

Figure 6. Bitcoin (Btc) price prediction results using LSTM, with training data shown in black, actual prices in blue, and predicted prices shown in red.

Result evaluation and comparison

The accuracy of the bitcoin price predictions generated by the various models was assessed using MAPE values. This metric was chosen for its computational simplicity and interpretation. MAPE values range from 0 to 100% (or 0 to 1) and are independent of the size of the variables being predicted. In general, the following classification applies when using MAPE for model evaluation (*45*):

- high predictive accuracy, 0–10% (0–0.1);
- good predictive accuracy, 10–20% (0.1–0.2);
- acceptable predictive accuracy, 20–50% (0.2–0.5).

The LSTM model achieved the lowest MAPE value of 5%, the VAR model showed a slightly higher error of 8% compared to the neural network model, while the Prophet model showed the lowest accuracy with a MAPE of 11% (**Table 5**).

Table 5. MAPE values for all models*

Model	MAPE
VAR	0.08 (8%)
Prophet	0.11 (11%)
LSTM	0.05 (5%)

*Abbreviations: LSTM – long short-term memory neural network, MAPE – mean absolute percentage error, VAR – vector autoregression.

Discussion

We achieved high accuracy in predicting bitcoin prices with VAR and LSTM models; while the Prophet model performed below either of the two, its results were still satisfactory, with the model having practical advantages due to its simplicity and time efficiency.

Although the LSTM model achieved high predictive accuracy, a 5% margin of error limits its applicability for real-world scenarios, especially for high-stakes or intra-day cryptocurrency trading. In general, LSTM models effectively captures non-linear dependencies and long-term temporal patterns (*46*), while the Prophet model focuses on trends and seasonality. However, previous research (*47*) did not find strong calendar effects in cryptocurrency returns, which may explain the superior performance of LSTM in modeling the relationship between social media sentiment and bitcoin prices in our study. Similar results reported by Boozary and colleagues (*46*) emphasize the LSTM models' strength in capturing complex price dynamics. Despite the limitations of our study, its results provide theoretical insights into bitcoin price formation and emphasize the predictive value of behavioral factors. Furthermore, the empirical results show that machine learning models such as the LSTM have lower error rates compared to the traditional VAR model. Importantly, the analysis confirms that sentiment expressed in Twitter posts is positively correlated with bitcoin price, with increasing sentiment polarity being associated with an increase in price.

The observed period reflects the characteristics of a speculative bubble, which is characterized by peak valuations and a subsequent decline. The results must therefore be interpreted against the backdrop of bitcoin's unprecedented price level. While this timeframe captures strong sentiment dynamics, the results may not fully generalize to more stable market conditions. Social media sentiment proves to be an important driver of price formation in such high price phases and provides valuable input for predictive models. The Granger causality between the polarity of sentiment and bitcoin highlights behavioral influences and the speculative nature of cryptocurrency investments. These findings suggest that the valuation of bitcoin is strongly influenced by subjective factors, providing important insights into its phenomenology, investment risks, and potential role as a legal tender.

Sentiment analysis that integrates multiple dimensions of information from tweets, such as the presence of positive or negative messages, the influence or credibility of the authors of the tweets, and the context of the messages, generally leads to better predictive results than relying only on tweet volume or the sheer number of tweets. For example, it has been reported that the volume of tweets has a measurable impact on the liquidity of the bitcoin market (*48*). Notably, tweet volume significantly influences liquidity only within the first 10 minutes after the tweets are posted, and this influence quickly dissipates and disappears within about 60 minutes. Therefore, sentiment indices that combine multiple attributes of social media posts provide a more robust input for modeling cryptocurrency price dynamics than simple volume-based metrics. In addition, while Google Trends can be used to create a sentiment index, prior research found them to have no long-term statistical impact on cryptocurrency prices (*49*).

Our study integrates social media sentiment along with additional behavioral indicators that reflect public interest in bitcoin. While much of the literature considers price prediction as a classification task, this paper uses regression models to predict absolute price values. The extended prediction horizon chosen here also represents a deviation from existing literature without compromising prediction accuracy. A major limitation of this analysis is its sample size; the final dataset contains a relatively small number of observations, which reduces the generalizability of our findings. Future research should favor longer time periods to capture periods beyond peak market conditions.

Provenance: Submitted. The article is based on the master's thesis by Dora Grubišić, deposited in the Dabar repository (https://urn.nsk.hr/urn:nbn:hr:124:091042). The thesis was defended at the Faculty of Economics of the University of Split in October 2023, under the supervision of Professor Mario Jadrić

Submitted: 09 December 2024 / **Accepted:** 23 June 2025 / **Published:** 31 October 2025

Peer review: Externally reviewed.

Availability of data: The raw data for this study are available upon request to the corresponding author.

Funding: No funding was received for this study.

Authorship declaration: DG participated in defining the topic of the study and carried out all experimental parts of the study and contributed to the collection, analysis, and interpretation of the data. BŠP and MJ developed the original research topic and contributed to the research design and the conceptualization and interpretation of the data. All authors wrote the first version of the manuscript and revised it for the final version. All authors agreed on the manuscript to be submitted and published and take full accountability for the content.

Disclosure of interest: The authors have completed the ICMJE Disclosure of Interest Form (available on request from the corresponding author) and disclose no relevant interests.

ORCID

Blanka Škrabić Perić https://orcid.org/0000-0002-7448-3821

Mario Jadrić https://orcid.org/0000-0002-2591-3899

References

1. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system [Internet]. Bitcoin.org; 2008 [cited 2023 Mar 02]. Available from: https://bitcoin.org/bitcoin.pdf.
2. European Central Bank. What is bitcoin? [Internet]. 2021 [cited 2023 Mar 02]. Available from: https://www.ecb.europa.eu/ecb/educational/explainers/tell-me/html/what-is-bitcoin.en.html.
3. Hrvatska narodna banka. Rizici povezani s kriptoimovinom [Internet]. 2021 [cited 2023 Mar 11]. Available from: https://www.hnb.hr/-/rizici-povezani-s-kriptoimovinom.
4. Söderberg G. Are Bitcoin and other crypto-assets money? [Internet]. Economic Commentaries No. 5, Sveriges Riksbank; 2018 [cited 2023 Mar 31]. Available from: https://www.riksbank.se/globalassets/media/rapporter/ekonomiska-kommentarer/engelska/2018/are-bitcoin-and-other-crypto-assets-money.pdf.

5. Vora G. Cryptocurrencies: are disruptive financial innovations here? Mod Econ. 2015;6:816–32. https://doi.org/10.4236/me.2015.67077

6. Kubat M. Virtual currency bitcoin in the scope of money definition and store of value. Procedia Econ Finance. 2015;30:409–16. https://doi.org/10.1016/S2212-5671(15)01308-8

7. Bindsell U, Schaaf J. Bitcoin's last stand [Internet]. European Central Bank blog; 2022 [cited 2023 Mar 31]. Available from: https://www.ecb.europa.eu/press/blog/date/2022/html/ecb.blog221130~5301eecd19.en.html.

8. Hackethal A, Hanspal T, Lammer DM, Rink K. The characteristics and portfolio behavior of bitcoin investors: evidence from indirect cryptocurrency investments. SSRN [preprint]. 2021 [cited 2023 Mar 31]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3501549.

9. Pelster M, Breitmayer B, Hasso T. Are cryptocurrency traders pioneers or just risk-seekers? Evidence from brokerage accounts. Econ Lett. 2019;182:98–101. https://doi.org/10.1016/j.econlet.2019.06.013

10. Nadler P, Guo Y. The fair value of a token: how do markets price cryptocurrencies? Res Int Bus Finance. 2019;52:101108. https://doi.org/10.1016/j.ribaf.2019.101108

11. Kristoufek L. What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. PLoS One. 2015;10(4):e0123923. https://doi.org/10.1371/journal.pone.0123923

12. Poyser O. Exploring the dynamics of bitcoin's price: a Bayesian structural time series approach. Eurasian Economic Review. 2019;9:29–60. https://doi.org/10.1007/s40822-018-0108-2

13. Buchholz M, Delaney J, Parker J, Warren J. Bits and bets: information, price volatility, and demand for bitcoin [internet]. reed.edu; 2012 [cited 2023 Mar 31]. Available from: https://www.reed.edu/economics/parker/s12/312/finalproj/Bitcoin.pdf

14. Ciaian P, Kancs A, Rajcaniova M. The economics of bitcoin price formation. Appl Econ. 2016;48(19):1799–815. https://doi.org/10.1080/00036846.2015.1109038

15. Zhu Y, Dickinson D, Li J. Analysis of the influence factors of bitcoin's price based on VEC model. Financ Innov. 2017;3:3. https://doi.org/10.1186/s40854-017-0054-0

16. Almeida J, Cruz Gonçalves T. A systematic literature review of investor behavior in the cryptocurrency markets. J Behav Exp Finance. 2023;37:100785. https://doi.org/10.1016/j.jbef.2022.100785

17. Nosfinger J. Social mood and financial economics. J Behav Finance. 2005;6(3):144–60. https://doi.org/10.1207/s15427579jpfm0603_4

18. Cambria E. Affective computing and sentiment analysis. IEEE Intell Syst. 2016;31(2):102–7. https://doi.org/10.1109/MIS.2016.31

19. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. J Comput Sci. 2011;2(1):1–8. https://doi.org/10.1016/j.jocs.2010.12.007

20. Bondo Hansen K, Borch C. Alternative data and sentiment analysis: prospecting non-standard data in machine learning-driven finance. Big Data Soc. 2022; (Jan-Jun):1–14.

21. Gottal A, Mittal A. Stock prediction using Twitter sentiment analysis. stanfrod.edu; 2011 [cited 2023 Mar 23]. Available from: https://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf.

22. Gupta A, Dengre V, Kheruwala HA, Shah M. Comprehensive review of text-mining applications in finance. Financ Innov. 2020;6:39. https://doi.org/10.1186/s40854-020-00205-1

23. Abraham J, Higdon D, Nelson J, Ibarra J. Cryptocurrency price prediction using tweet volumes and sentiment analysis. SMU Data Sci Rev. 2018;1(3):1.

24. Algamdi S, Alhakami H, Alqethami S, Alsubait T. Cryptocurrency price prediction using forecasting and sentiment analysis. Int J Adv Comput Sci Appl. 2022;13(10) https://doi.org/10.14569/IJACSA.2022.01310105

25. Ellul J, Gatt A, Vella Critien J. Bitcoin price change and trend prediction through Twitter sentiment and data volume. Financ Innov. 2022;8:1–19.

26. Lamon C, Nielsen E, Redondo E. Cryptocurrency price prediction using news and social media sentiment. SMU Data Sci Rev. 2017;1(3):1–22.

27. Valencia F, Gómez-Espinoza A, Valdés-Aguirre B. Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. Entropy (Basel). 2019;21(6):589. https://doi.org/10.3390/e21060589

28. Gurrib I, Kamalov F. Predicting bitcoin price movements using sentiment analysis: a machine learning approach. Stud Econ Finance. 2022;39(3):347–64. https://doi.org/10.1108/SEF-07-2021-0293

29. Wong EXL. Prediction of bitcoin prices using Twitter data and natural language processing [Internet]. duke.edu; 2021 [cited 2023 Mar 31]. Available from: https://hdl.handle.net/10161/24081.

30. Fung K, Jeong J, Pereira J. More to cryptos than bitcoin: a GARCH modelling of heterogeneous cryptocurrencies. Finance Res Lett. 2022;47(Pt A):102544. https://doi.org/10.1016/j.frl.2021.102544

31. Phung DQ, Nguyen TO, Le HP, Pham HH, Luong KL, Nguyen NK. Estimating and forecasting bitcoin daily prices using ARIMA-GARCH models. Bus Anal J. 2024;45(1):11–23. https://doi.org/10.1108/BAJ-05-2024-0027

32. Saeedi A, Al Fattal A. Examining trust in cryptocurrency investment: insights from the structural equation modeling. Technol Forecast Soc Change. 2024;210:123882. https://doi.org/10.1016/j.techfore.2024.123882

33. Škrinjarić T. Odabrane teme primijenjene ekonometrije: uvod u analizu vremenskih nizova. Zagreb (Croatia): Hrvatska narodna banka; 2023.

34. JustAnotherArchivist. snscrape: a social networking service scraper in Python [Internet]. github.com; 2023 [cited 2023 Jan 18]. Available from: https://github.com/JustAnotherArchivist/snscrape.

35. Gilbert EE, Hutto CJ. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media; 2014 Jun 1–4; Ann Arbor, Michigan, USA. Washington D.C.: Association for the Advancement of Artificial Intelligence; 2014. p. 216–225.

36. Bitcoing – Google Trends [Internet]. trends.google.com; 2023 [cited 2023 Feb 06]. Available from: https://trends.google.com/trends/explore?date=2021-10-31%202021-12-30&q=bitcoin&hl=en.

37. Pageviews Analysis [Internet]. pageviews.wmcolud; 2023 [cited 2023 Feb 06]. Available from: https://pageviews.wmcloud.org/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2021-10-31&end=2021-12-30&pages=Bitcoin.

38. Bitcoin USD. (BTC-USD) [Internet]. finance.yahoo; 2023 [cited 2023 Feb 06]. Available from: https://finance.yahoo.com/quote/BTC-USD/history?period1=1635638400&period2=1640822400&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true.

39. Oikonomopoulos S, Tzafilkou K, Karapiperis D, Verykios V. Cryptocurrency price prediction using social media sentiment analysis. In: Proceedings of the 13th International Conference on Information, Intelligence, Systems & Applications (IISA); 2022 Jul 18–20; Corfu, Greece. New York (NY): IEEE; 2022. p. 1–8.

40. Cheng J, Tiwari S, Khaled D, Mahendru M, Shahzad U. Forecasting bitcoin prices using artificial intelligence: combination of ML, SARIMA, and Facebook Prophet models. Technol Forecast Soc Change. 2024;198:122938. https://doi.org/10.1016/j.techfore.2023.122938

41. Ozcicek O, McMillin D. Lag length selection in vector autoregressive models: symmetric and asymmetric lags. Appl Econ. 1999;31(4):517–24. https://doi.org/10.1080/000368499324237

42. Prophet [Internet]. facebook.github.io; 2023 [cited 2023 Apr 03]. Available from: https://facebook.github.io/prophet/.

43. Peixeiro M. Time series forecasting in Python. Shelter Island (NY): Manning Publications; 2022.

44. Terven J, Cordova-Esparza D, Ramirez-Pedraza A, Chávez Urbiola E. Loss functions and metrics in deep learning: a review. arXiv:2307.02496 [preprint]. Available from: https://arxiv.org/abs/2307.02694

45. Lewis CD. Industrial and business forecasting methods. London (UK): Buttersworth Publishing; 1982.

46. Boozary P, Sheykhan S. GhorbanTanhaei H. Forecasting the bitcoin price using various machine learning techniques: a systematic review in data-driven marketing. Syst Soft Comput. 2025;7:200209. https://doi.org/10.1016/j.sasc.2025.200209

47. Kaiser L. Seasonality in cryptocurrencies. Finance Res Lett. 2019;31:100386. https://doi.org/10.1016/j.frl.2018.11.007

48. Choi H. Investor attention and bitcoin liquidity: evidence from bitcoin tweets. Finance Res Lett. 2021;39:101555. https://doi.org/10.1016/j.frl.2020.101555

49. Bonaparte Y, Bernile G. A new "Wall Street darling?" Effects of regulation sentiment in cryptocurrency markets. Finance Res Lett. 2023;52:103376. https://doi.org/10.1016/j.frl.2022.103376

RESEARCH ARTICLE