Asia-Pacific Financial Markets
https://doi.org/10.1007/s10690-024-09510-6

ORIGINAL RESEARCH

Crypto Volatility Forecasting: Mounting a HAR, Sentiment, and Machine Learning Horserace

Alexander Brauneis[1] · Mehmet Sahiner[2]

Accepted: 23 November 2024

Abstract

The relationship between investor sentiment and cryptocurrency market volatility remains an area of growing interest in empirical finance. In this study, we present an innovative forecasting approach by utilizing a unique dataset of AI-generated sentiment from a comprehensive database of crypto market news. In a horserace fashion, we first evaluate the Heterogeneous Autoregressive (HAR) model and then compare its forecasting performance to five advanced machine learning (ML) methods. ML performs reasonably well and improves the accuracy of the benchmark HAR model. Interestingly, including sentiment does not improve the forecasting accuracy of the HAR model. However, our findings highlight that investor sentiment seems to influence crypto market volatility in a nonlinear fashion that can (only) be captured by ML methods. In other words, LightGBM, XGBoost, and LSTM models show enhanced predictive accuracy when sentiment data is incorporated, improving no-sentiment forecasts in 54.17% of the cases studied. Overall, our results emphasize the significant potential of integrating machine learning and sentiment analysis as a promising avenue for improved forecasting, offering potential benefits for risk management strategies and provide valuable insights for researchers and practitioners.

Keywords Cryptocurrencies · Sentiment · Machine learning · Volatility forecasting

JEL Classification G10 · G14 · G17

✉ Alexander Brauneis
alexander.brauneis@ntu.ac.uk
https://www.ntu.ac.uk/staff-profiles/business/alexander-brauneis

Mehmet Sahiner
msahiner001@dundee.ac.uk
https://www.dundee.ac.uk/people/mehmet-sahiner

1 Finance and Accounting, Nottingham Trent University Business School, 50 Shakespeare Street, Nottingham NG1 4FQ, UK

2 Economics, School of Business, University of Dundee, Nethergate, Dundee DD1 4HN, Scotland, UK

Published online: 13 December 2024

1 Introduction

Cryptocurrencies have received immense interest from investors and policymakers due to their decentralized and traceable features (Fry and Cheah (2016)), while at the same time their distinctly volatile nature has prompted researchers to focus on predicting price movements by applying various techniques Köchling et al. (2020); Ftiti et al. (2021); Catania and Grassi (2022a); Xia et al. (2023). The majority of related works have utilized different econometric models to capture the intricate dynamics of cryptocurrency prices, such as the GARCH (Katsiampa (2017)) and HAR Qiu et al. (2021) models. Only some studies though have explored more advanced methods, such as machine learning (ML) Akyildirim et al. (2021) and deep learning (DL) Liu et al. (2021).

The growing body of research has also attempted to model the sharp movements in cryptocurrency prices by using various explanatory variables such as macro events, technical indicators, and social media indicators Wang et al. (xxxx); Hoang and Baur (2020); Kyriazis et al. (2022); Moser and Brauneis (2023). For example, Rognone et al. (2020) found that the movements in cryptocurrencies are mainly driven by investor enthusiasm, regardless of the market news direction.

In this context of crypto volatility forecasting, a comprehensive review of related studies (see Table 1 for details) focusing on traditional methods like GARCH and HAR, more advanced methods from the realm of machine learning and deep learning, as well as sentiment analysis reveals that as of yet no paper has addressed a comparative study of all these elements. In other words, there is respective and separate evidence for the suitability of e.g. particular GARCH specifications, machine learning approaches and the predictive power of sentiment. However, to the best of our knowledge, no study has yet combined both traditional and AI methods with sentiment data. We aim at filling this research gap by providing evidence on forecasting accuracy using the traditional HAR model, a variety of machine learning tools as well as crypto market sentiment.

Furthermore, and in particular, there is still a limited understanding of how sentiment analysis of economic and financial news relates to cryptocurrencies. While investor sentiment derived from media coverage as well as search volume is an established predictor for traditional financial assets such as equities (Tetlock (2007), Da et al. (2014)) for almost two decades, it entered research papers on cryptocurrencies only sporadically. Although significant strides have been made in the realm of cryptocurrency forecasting in the recent years, challenges persist, particularly with regard to the lack of a methodological basis in sentiment analysis for these digital assets. Therefore, more research is needed to develop accurate and robust forecasting models that can capture the unique characteristics of cryptocurrencies.

The use of sentiment analysis to predict future movements in cryptocurrency prices, particularly for leading assets such as bitcoin and ethereum, is an essential area of study in today's financial landscape. Therefore, in this paper, we investigate the impact of news sentiment on cryptocurrency volatility by applying five well-established machine learning models, namely, Multi-layer Perceptron

Table 1 Literature review. This table summarizes the current state of literature on cryptocurrency volatility forecasting. We report the authors of the respective study (Authors), the resolution of the data in the empirical study (Frequency), the sample period (Sample Period), the cryptocurrencies under consideration in the empirical study (Coins), as well as if the empirical study draws on traditional forecasting methods such as GARCH (column Trad.), on artificial intelligence based methods including machine learning (column AI), and on sentiment data (column Sent.). No study except ours utilizes all three approaches in the attempt to forecast crypto volatility

No	Authors	Frequency	Sample Period	Coins	Trad	AI	Sent	ALL
1	Gradojevic et al. (2023)	Intra-day	11/5/2015–7/3/2019	Bitcoin		✓	✓	
2	Kraaijeveld and De Smedt (2020)	Intra-day and Daily	4/6/2018–4/8/2018	Bitcoin, Ethereum, Ripple, Litecoin, Bitcoin Cash, Cardano, Stellar, EOS, and Tron	✓		✓	
3	Akyildirim et al. (2021)	Intra-day	2/1/2020–10/9/2020	Bitcoin		✓		
4	Ftiti et al. (2021)	Intra-day	4/2018–6/2020	Bitcoin, Ethereum, Ethereum Classic, and Ripple	✓			
5	Atsalakis et al. (2019)	Daily	13/9/2011–12/10/2017	Bitcoin		✓		
6	Sun et al. (2020)	Daily	1/1/2018–30/6/2018	Top 42 cryptocurrencies		✓		
7	Liu et al. (2021)	Daily	7/2013–12/2019	Bitcoin		✓		
8	Catania et al. (2019)	Daily	8/8/2015–28/12/2017	Bitcoin, Ethereum, Ripple, Litecoin	✓			
9	Köchling et al. (2020)	Intra-day and Daily	30/11/2015–20/8/2018	Bitcoin	✓			
10	Catania and Grassi (2022b)	Daily	5/2013–9/2019	Bitcoin, Ethereum, Ripple, Litecoin	✓			
11	Jana et al. (2021)	Daily	10/1/2013–23/2/2019	Bitcoin		✓		
12	Trucís (2019)	Daily	13/9/2011 to 31/12/2017	Bitcoin	✓			
13	Walther et al. (2019)	Daily	1/5/2015–31/7/2019	Bitcoin, Ethereum, Litecoin, Stellar, and Ripple	✓			
14	Ángeles López-Cabarcos et al. (2021)	Daily	4/1/2016–30/9/2019	Bitcoin	✓		✓	
15	Bouteska et al. (2022)	Daily	1/1/2015–30/9/2020	Bitcoin	✓		✓	
16	Katsiampa (2017)	Daily	18/7/2010–1/10/2016	Bitcoin	✓			
17	Balcilar et al. (2017)	Daily	9/12/2011–25/4/2016	Bitcoin	✓			
18	Bouri and Gupta (2021)	Monthly	7/2010–5/2019	Bitcoin	✓		✓	
19	Corbet et al. (2020)	Daily	19/7/2010–30/9/2019	Bitcoin	✓		✓	

Table 1 (continued)

No	Authors	Frequency	Sample Period	Coins	Trad	AI	Sent	ALL
20	Kumar Kulbhaskar and Subramaniam (2023)	Daily	4/2013–6/2021	Bitcoin, Ethereum, Ripple, Litecoin, Bitcoin Cash, Dash, EOS, Zcash, Monero, and Ethereum Classic	✓		✓	
21	This study	Intra-day	12/2021–12/2022	Bitcoin, Ethereum, Cardano, Dogecoin, Polkadot, Litecoin, Shiba Inu, Solana	✓	✓	✓	✓

(MLP), Long-Short Term Memory (LSTM), Extreme Gradient Boost Decision (XGBoost), Light Gradient Boosted Machine (LightGBM), and the hybrid CNN-bidirectional LSTM (CNNBiLSTM) in addition to the standard HAR model. We address the question whether news sentiment carries informational content and can be utilized to improve the accuracy of volatility forecasts in intraday data. In a nutshell, we specifically tackle two questions within a horserace-like framework. Firstly, do machine learning techniques provide more accurate forecasts of crypto volatility than the established HAR model? Secondly, does the inclusion of market sentiment as an explanatory variable enhance the accuracy of forecasts of either model?

Our results show that the machine learning models outperform traditional volatility forecasting methods (i.e.,the HAR model), indicating the efficacy of machine learning in predicting cryptocurrency market movements. In particular, we find that deep learning and LightGBM classifiers are well-suited for this forecasting problem. Furthermore, sentiment seems to be a predictor of cryptocurrency volatility in the sense of more accurate forecasts, respective results are ambiguous however. For three out of six candidate models, the inclusion of sentiment substantially improves forecasting accuracy, while for the other models forecasts deteriorate.

Our research makes a novel contribution to the literature by comparing the predictive capability of various machine learning models, such as gradient boosting models, artificial neural networks, and deep learning, along with different feature sets, including market sentiment. Our study establishes a comprehensive benchmark for the accuracy of short-term cryptocurrency volatility forecasting models.

The remainder of the paper is organized as follows. In section 3 we describe the data used in the empirical analyses, section 4 outlines the methodology. The results are presented in section 5, and section 6 concludes.

2 Literature review

The importance of sentiment analysis in financial markets, particularly within the dynamic and volatile domain of cryptocurrencies, has gained prominence in the recent years. Early studies in sentiment analysis, such as those by Gilbert and Karahalios (2010), focused on general market sentiment derived from social media platforms and its correlation with stock market indices. In a similar vein, Bollen et al. (2011) demonstrated capabilities of social media sentiment, notably on stock market movements, through the analysis of data from X (formerly known as Twitter). This emerging field has been further expanded in the context of cryptocurrencies, a relatively new asset class known for its volatility and sensitivity to market sentiment. A growing body of literature have investigated how sentiment affects cryptocurrency valuations using different proxies. These methods range from analysing Google search trends (Panagiotidis et al. (2019)) and search data from Wikipedia (Kristoufek (2013)) to assessing sentiment in news articles (Corbet et al. (2020)), opinions expressed on X (Shen et al. (2019)), and discussions in cryptocurrency blogs (Karalevicius et al. (2018)). However, the extension of these findings in the cryptocurrency market show that it present significant challenges due to the unique market dynamics and the intricate nature of digital assets, as highlighted by studies like Guégan and Renault (2021) which has been exploring the impact of investor sentiment on cryptocurrency prices. Despite these efforts, more comprehensive analytical scope is required to fully comprehend the complexities of cryptocurrency sentiment dynamics, highlighting a distinct gap in the existing literature (Peng et al. (2023)).

Advancements in machine learning (ML) models have opened new possibilities for financial market analysis, with their application in cryptocurrencies being a more recent phenomenon. Specifically, the application of sophisticated ML algorithms, ranging from established methods like Random Forests to more advanced techniques like Neural Networks, has demonstrated improved prediction accuracy over conventional econometric models. The initial applications, exemplified by McNally et al. (2018), showed the potential of simple machine learning models in predicting cryptocurrency prices but often overlooked the crucial interplay between market sentiment and price volatility. The more advanced LSTM networks, as explored by Alessandretti et al. (2018), marked a significant step towards incorporating complex market data. Similarly, Mallqui and Fernandes (2019) further demonstrated the effectiveness of ML models in predicting cryptocurrency prices, yet their approach lacked thoroughness in data selection and feature engineering. However, the integration of sentiment analysis with these advanced machine learning models, especially utilizing comprehensive sentiment data for enhancing prediction accuracy, remains relatively untapped.

In the area of cryptocurrency sentiment analysis with machine learning, a pivotal study conducted by Bukovina and Marticek (2016) stands as a key study. They examined the influence of investor sentiment on Bitcoin volatility, utilizing the autoregressive AR(1) model and the Sentdex sentiment index, which applies Natural Language Processing (NLP) techniques for sentiment extraction from textual data. NLP, particularly in a sentiment analysis, is a widely used approach for interpreting opinions from various sources such as customer reviews, surveys, and social media comments. The findings of their study revealed that while sentiment only explained a minor portion of total volatility, its impact was more pronounced during periods of high volatility. Extending this approach, Abraham et al. (2018) demonstrated the effectiveness of this combination in forecasting cryptocurrency prices. Their work highlights the incremental benefits of incorporating sentiment data into machine learning models, suggesting a more sophisticated understanding of market dynamics than what is achievable by either approach alone. Figá-Talamanca and Patacca (2019) also explored the impact of investor sentiment on cryptocurrency returns, using ARMA, GARCH, and EGARCH models. They employed trading volume and Google search index as proxies for investor sentiment. Their results indicated that trading volume affected both the mean and volatility of cryptocurrency returns, while the search index primarily influenced volatility.

In more recent studies, such as those by Naeem et al. (2021), and Bouri et al. (2021), the focus has expanded to encompass a broader range of cryptocurrencies and more diverse sentiment measures. Naeem et al. (2021) used the FEARS index and Twitter happiness sentiment to analyze the impact on returns of six major cryptocurrencies. They found that while the happiness sentiment index significantly predicted returns, the FEARS index showed weaker and short-term predictability. Bouri et al. (2021) took a different approach, examining volatility spillover across fifteen major cryptocurrencies by considering the impact of investor sentiment derived from Twitter feeds. Their findings indicated that extreme unhappiness among investors led to increased market volatility and interconnectedness, whereas extreme happiness correlated with lower total connectedness, suggesting potential diversification opportunities between cryptocurrencies. Another dimension was added by Plakandaras et al. (2021), who investigated the potential role of the US-China Trade War as a sentiment variable. Using machine learning techniques and regression models, they assessed the trade war's impact on Bitcoin returns, finding weak evidence to suggest that trade-related uncertainties significantly influenced Bitcoin's future returns.

These studies collectively underscore the importance of investor sentiment in understanding the return and volatility dynamics of cryptocurrency market. The full potential of advanced techniques such as ensemble methods and deep learning architectures, in conjunction with nuanced sentiment analysis, remains largely unexplored. This integration is crucial for addressing the non-linearities and complexities inherent in cryptocurrency markets, which have been frequently oversimplified in existing models. By integrating advanced sentiment data with sophisticated machine learning models against the traditional Heterogeneous Autoregressive (HAR) model, our research aims to bridge this methodological gap in the existing literature. We offer a novel perspective on the impact of investor sentiment on volatility of different cryptocurrencies, adding a critical layer to our understanding of cryptocurrency market dynamics. This approach is particularly relevant in the current context, where traditional models may not fully capture the nuances of market behaviour of less pronounced cryptocurrencies. Our study not only contributes to the theoretical framework around cryptocurrency market analysis but also provides practical insights for investors and market analysts. This research, therefore, stands as a significant addition to the existing body of literature, offering a comprehensive and timely analysis of cryptocurrency market dynamics in the face of global challenges.

3 Data

The dataset in this study includes 8 major cryptocurrencies traded against the USD on the leading crypto trading venue Coinbase. These coins are bitcoin, ethereum, polkadot, shiba inu, solana, cardano, doge coin and litecoin. The time period under consideration spans from December 21, 2021 00:00 UTC to December 22, 2022 06:00 UTC. During this period, the coins in our sample accounted for almost 70% of the total cryptocurrency market capitalization. We retrieve price data for each coin from the Coinbase API on a 5-minute resolution, which is then used to calculate realized volatilities. Motivated by the findings of Liu et al. (2015), we opt for this volatility measure as the focal variable in this analysis.

Furthermore, we collect crypto market news data from headlinehunter.ai, a commercial platform that acquires online content from a vast number of websites including traditional news such as tipranks.com, bloomberg.com and forbes.com, crypto-specific news from e.g. cointelegraph.com and decrypt.co as well as social media such as X (twitter) and reddit. Non-english sources are being translated, subsequently, headlinhunter.ai applies a generative artificial intelligence based computer linguistic model to the text which generates a sentiment score from +1 to -1 (positive/neutral/negative, as well as: no sentiment) for every single text. Additionally, all online contents are classified, i.e., assigned to different news channels. These are: coin specific news (i.e., explicitly referring to bitcoin, ethereum,...), general crypto news, news related to crypto mining, news on regulation and legislation issues with cryptocurrencies, content created by 'influencers' (e.g., popular and verified twitter accounts), as well as news related to Covid 19 which may also be a driver of crypto volatility given the time period of the dataset. For each news channel, we finally retrieve a time series of sentiment data on a 6-hours resolution. This includes (1) the total number of retrieved online contents attributed to a channel, (2) the average sentiment of these news and (3) the density of news, defined as the number of non no-sentiment news divided by the total number of news. Hence, we may capture the extent of news coming to the market, the general sentiment, as well as the resilience of the sentiment proxied by the fraction in (3). While headlinehunter.ai certainly does not capture crypto market online content in its entirety, we argue that it provides a representative sample of news.

This analysis uses 6-hours intervals of data with a total of 1,461 observations for each time series for two reasons. First, while some news channels (e.g., crypto news) show hundreds of news within a 6-hours interval on average, other more specific channels (e.g., crypto mining) occasionally feature only a low two digit number of news within 6 h. Therefore, shortening the time interval to say one hour would result in an insignificant amount of news for some channels. Second, extending the time interval to say one day would result in insufficient observations for the data-intensive ML models, given the sample period of one year only.

4 Methodology

The main goal is to one-step ahead forecast realized volatility (RV) of cryptocurrencies. RV are calculated for 6-hours intervals t according to:

$$RV_t = \sum_{i=2}^{72} \log \left(\frac{P_{ti}}{P_{ti-1}} \right)^2 \tag{1}$$

where P_{ti} denotes the close price of the 5-minute subinterval i in the 6-hours period t. Table 2 reports descriptive statistics of RV for each of the 8 coins in the sample respectively.

Table 2 **Descriptive statistics for 6-hour realized volatilities** for the eight cryptocurrencies in the sample (bitcoin-BTC, ethereum-ETH, cardano-ADA, doge coin-DOGE, polkadot-DOT, litecoin-LTC, shiba inu-SHIB, solana-SOL). The table reports mean, median, the 25% and 75% quantile (q25, q75), standard deviation (Std. Dev.), all values in percent, as well as skewness and kurtosis of realized volatilities. The last two rows report p-values of the Jarque-Bera (JB) and the augmented Dickey-Fuller (ADF) test truncated at 0.001, rejecting the normality and unit-root null

Statistic	BTC	ETH	ADA	DOGE	DOT	LTC	SHIB	SOL
Mean	1.49	1.99	2.29	2.51	2.37	2.17	3.41	2.83
Median	1.29	1.73	1.95	2.00	2.10	1.91	3.14	2.40
q25	0.93	1.25	1.49	1.47	1.59	1.44	2.41	1.79
q75	1.76	2.35	2.69	2.89	2.79	2.52	4.01	3.26
Std. Dev	0.92	1.18	1.34	1.76	1.38	1.19	1.70	2.00
Skewness	2.33	2.42	2.71	3.35	3.66	2.92	2.49	4.78
Kurtosis	11.04	12.02	15.68	21.54	32.93	17.67	15.52	40.75
JB	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
ADF	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

4.1 Forecasting models

The heterogeneous autoregressive (HAR) model has proven to work well for cryptocurrencies (cf. Qiu et al. (2021)). This model proposed by Corsi (2009) uses lagged RV as well as past weekly and monthly averages as predictors for RV and is defined as follows:

$$RV_t = c + \beta_1 \cdot RV_{t-1} + \beta_2 RV_{t-1}^{(8)} + \beta_3 RV_{t-1}^{(28)} + \vec{\gamma} \cdot \mathbf{X}_{t-1} + \epsilon_t \tag{2}$$

where $RV_t^{(8)}$ ($RV_t^{(28)}$) denotes the average realized volatility for 8 (28) periods prior to and including t, corresponding to 2 and 7 days respectively with our data frequency. We deviate from the HAR convention of averaging over 5 and 21 prior (daily) observations (a trading week / month in traditional markets) for two reasons. First, crypto markets operate 24/7/365, therefore a averaging over 7 (30) periods seems more appropriate. Second, the volatility of cryptocurrencies features intradaily patterns (e.g., see Brauneis et al. (2024)), we thus choose to average over a multiple of four 6-hours periods, corresponding to one full day. Finally, ϵ is the error term and $\mathbf{X}$ is the set of explanatory sentiment variables including the total number of news, their average sentiment as well as the density.

In our pursuit of determining the most accurate forecasting model, we address the question of whether using complex machine learning methods consistently outperform simpler techniques such as the HAR model. We consider five distinct machine learning algorithms: LightGBM, XGBoost, LSTM, MLP, and CNNBiLSTM models against the benchmark model of HAR due to their demonstrated efficacy in addressing the challenges posed by the rapid changes and uncertainties inherent to cryptocurrency markets, as highlighted by McNally et al. (2018) and Akyildirim et al. (2021). We specifically select the gradient boosting frameworks of LightGBM and XGBoost for their ability to capture complex patterns and insights from sentiment data (Krauss et al. (2017)). Moreover, prior research has also demonstrated the superiority of ANNs (Sermpinis et al. (2021)) and LSTM models (Chen et al. (2021)) over traditional prediction methods. Therefore, by leveraging the strengths of these established models, our aim is to effectively capture the expressive power of sentiment dynamics and enhance the accuracy of volatility prediction in cryptocurrency market. The fitting and evaluation framework for XGBoost and LightGBM models differs from ANN and Deep Learning.

Chen and Guestrin (2016) introduced XGBoost as a method to address both classification and regression tasks by optimizing the objective function through the incorporation of regularization terms and increasing the prediction accuracy using the second-order Taylor expansion (Wei et al. (2023)). On the other hand, Ke et al. (2017) developed LightGBM to address computational inefficiencies in large datasets without significant compromise on accuracy. When compared to XGBoost, LightGBM exhibits two primary advantages regarding time efficiency which are the selection of the optimal splitting node and its distinct tree growth approach.

Initially, we employ a structured approach involving training, validation, and test datasets. A 5-fold cross-validation approach is then utilized to fit each model on corresponding validation sets with fixed hyperparameter values. This process is repeated under five iterations, and the average performance metrics are used to assess the models' predictive capabilities. Hyperparameters are optimized using a similar cross-validation process, and the final chosen values are used to evaluate the models on the holdout test dataset. It is important to note that the test data is excluded from the training phase and can be considered as a novel future data for the trained models. In contrast, the training and testing procedures for MLP, LSTM and CNN-BiLSTM are calculated based on the Bayesian Optimization technique of Wu et al. (2019) and the optimized hyperparameters are implemented for each model using TensorFlow 2.0. For further details see Cho et al. (2020) and Sahiner et al. (2023).

With this set of models, we generate one-step ahead forecasts of realized volatility, respectively. 80% of the total number of observations, i.e. 1,169 out of 1,461 are used to train the machine learning models. Specifically, ML models are trained on RV_{t-1}, $RV_{t-1}^{(8)}$, $RV_{t-1}^{(28)}$, and $\mathbf{X}_{t-1}$ and the known outcome RV_t. At the same time, we estimate coefficients of the HAR model in Eq. 2 based on 1,169 data points. Subsequently, one-step ahead forecasts are derived from HAR and ML models respectively. We repeat this process by employing a moving window of input data until the end of the dataset, resulting in a total of 292 forecasts of RV. Figure 1 presents a flowchart of the empirical approach.

4.2 ML algorithms

In this section we briefly outline the machine learning algorithms employed in the empirical analyses. The objective of this is to highlight the different conceptual approaches of the ML tools, which possibly entail diverging outcomes in terms of their ability to forecast cryptocurrency realized volatility.

LightGBM Model

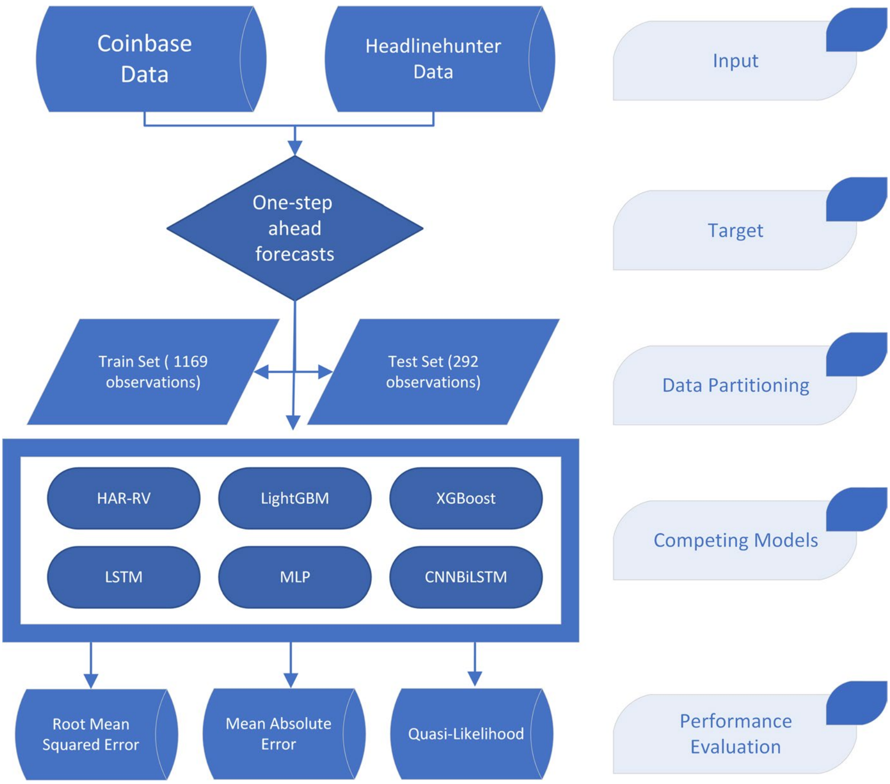

Fig. 1 **Methodology flowchart.** This figure outlines the structure of the empirical analysis, based on price data from Coinbase and sentiment data. All competing models are trained / calibrated with a total of 1,169 observations from each explanatory variable (Train Set). Then, we derive a total of 292 one-step ahead forecasts (Test Set) using a rolling window of input data. Generally, the t^{th} forecast relies on observations $\{t, t+1, ..., t+1168\}$ from each time series. Finally, we use three loss functions to assess the accuracy of the predictions

LightGBM is a relatively new algorithm within the Gradient Boosting Decision Tree (GBDT) family, introduced by Ke et al. (2017). It has gained popularity due to its ability to handle large-scale data efficiently, making it particularly well-suited for tasks such as classification, regression, and prediction. Similar to XGBoost, LightGBM builds an ensemble of decision trees, but it introduces two key innovations: Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), which are crucial for optimizing model performance when dealing with high-dimensional datasets, such as cryptocurrency data.

The LightGBM model aims to approximate the function $f^*(x)$ that minimizes the expected loss over the training set $X = \{(x_i, y_1)\}_{i=1}^{n}$:

$$f = argmin\mathbf{E}_{X,y}[L(y, f(x))] \tag{3}$$

To achieve this, LightGBM integrates T regression trees in the form of:

$$f_T(X) = \sum_{t=1}^{T} f_t(X) \tag{4}$$

Each regression tree can be represented as $w_{q(x)}$, where $q \in \{1, 2, ..., J\}$ represents the decision rules of the tree and w is a vector that corresponds to the sample weights of the leaf nodes. LightGBM uses an additive learning strategy, where the model is updated iteratively, minimizing the loss function at each step:

$$t = \sum_{i=1}^{n} L(y_i, F_{t-1}(x_i) + f_t(x_i)) \tag{5}$$

In LightGBM, the objective function is quickly approximated using Newton's method. Simplifying the expression and removing constant terms results in the following form:

$$\sum_{i=1}^{n} (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) \tag{6}$$

where g_i and h_i represent the first- and second-order gradient statistics of the loss function, respectively. For a specific tree structure $q(x)$, the optimal weights w_j^* of the leaf nodes and the associated value Γ_K can be calculated as:

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{7}$$

where, I_j represents the sample set for leaf j, and λ is a regularization parameter. The scoring function T^*, which evaluates the quality of the tree structure q, is derived as follows:

$$T^* = \frac{1}{2} \sum_{j=1}^{J} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} \tag{8}$$

To improve the tree structure, LightGBM optimizes the split by maximizing the gain G, defined as:

$$G = \frac{1}{2} \left(\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right) \tag{9}$$

where, I_L and I_R represent the sample sets for the left and right branches of the tree, respectively. Unlike traditional GBDT-based algorithms, which grow trees horizontally, LightGBM builds trees vertically. This vertical growth strategy allows LightGBM to handle large-scale datasets and numerous features more efficiently.

XGBoost

XGBoost is one of the machine learning models designed to optimize both speed and performance by using decision tree ensembles. In XGBoost, each internal node of the decision tree represents a test of a feature, while the leaf nodes assign a score to the input data. The final prediction is obtained by summing the contributions of multiple decision trees, where the total output is the sum of the scores from K trees, expressed as:

$$y = \sum_{k=1}^{K} f_k(x_t) \tag{10}$$

where x_t represents the input features at time t (such as price, trading volume, or sentiment data), f_k is the score from the k-th regression tree, and K is the total number of trees used in the model. XGBoost operates similarly to other gradient boosting methods but introduces significant enhancements to the model's objective function, particularly by adding a regularization term that helps control the complexity of the model and reduces the risk of overfitting.

The regularized objective function in XGBoost can be written as:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{11}$$

where $l(y_i, \hat{y}_i)$ is the loss function that measures the difference between the predicted volatility $\hat{y}_i$ and observed volatility y_i. The regularization term $\Omega(f_k)$ penalizes the complexity of each tree and is defined as:

$$\Omega(f) = \phi T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{12}$$

where T represents the number of leaf nodes in the tree, w_j denotes the weight assigned to each leaf, and ϕ and λ are regularization parameters controlling the model's complexity. By including this regularization, XGBoost is better equipped to prevent overfitting, a common concern in financial markets where models can easily capture noise instead of meaningful patterns.

Multi-Layer Perceptron (MLP)

A Multi-Layer Perceptron (MLP) is a type of feedforward neural network widely employed for tasks such as pattern recognition, prediction, and approximation. Within this network, the connections between neurons are determined by weights, which are iteratively adjusted during the learning process. Training of MLP utilizes the backpropagation algorithm, which operates in two stages: a forward pass and a backward pass. In the forward pass, input data is processed through the network with fixed weights to generate an output. In the backward pass, the error between the actual and predicted outputs is computed and propagated backward through the network to adjust the weights. This iterative adjustment aims to minimize the mean or total squared error.

The MLP model with a single hidden layer is mathematically represented as follows:

$$n_{k,t} = w_{k,0} + \sum_{i=1}^{I} w_{k,i} x_{i,t} \tag{13}$$

$$N_{k,t} = L(n_{k,t}) = \frac{1}{1 + e^{-n_{k,t}}} \tag{14}$$

$$Y_t = \lambda_0 + \sum_{k=1}^{K} \lambda_k N_{k,t} \tag{15}$$

The training process begins with input vectors $x_{i,t}$, weight vectors $w_{k,i}$, and bias terms $w_{k,0}$. These inputs are passed through the logistic sigmoid function which is defined by $L(n_{k,t})$ as $\frac{1}{1+e^{-n_{k,t}}}$ to form neurons $N_{k,t}$ which in turn predict the output Y_t.

LSTM and CNN-BiLSTM

LSTM (Long Short-Term Memory) is a popular deep learning architecture within Recurrent Neural Networks (RNN) designed for time series prediction tasks. It has proven effective in both classification and regression problems across various domains, such as stock market and cryptocurrency prediction (Chen et al. (2021); Lin et al. (2021)). Although traditional RNNs (Recurrent Neural Network) can preserve information better than standard networks, they struggle to learn long-term dependencies due to the vanishing gradient problem (Hochreiter (1998)). LSTM addresses this issue by incorporating memory cells, which allow it to effectively manage long-term dependencies (Gers et al. (2000)).

The core component of the LSTM architecture is the cell state, which maintains information flow through the network with minimal changes, thereby preventing the vanishing gradient issue. The LSTM model utilizes a gate mechanism comprising three gates: the input gate, forget gate, and output gate. These gates selectively manage the flow of information through the network. The forget gate determines which information from the previous cell state should be discarded, while the input gate updates the cell state with new information. Finally, the output gate decides which information from the current cell state should be passed to the hidden state.

The mathematical formulation of LSTM is represented as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{16}$$

$$\tilde{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_f) \tag{17}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{18}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{19}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{20}$$

$$h_t = o_t \odot tanh(c_t) \tag{21}$$

In the equations, f_t, i_t, o_t, c_t, and $\tilde{c}_t$ represent the forget gate, input gate, output gate, cell state and temporary state of the cell, respectively. The matrices W_f, W_i, and W_o correspond to the weight coefficients associated with the forget, input, and output gates. Similarly, b_f, b_i, and b_0 are the bias terms corresponding to each of these gates. The variable U denotes the internal state of the neuron, while h_t represents the output from the hidden layer. The function σ refers to the Sigmoid activation function, and the symbol $\odot$ indicates the Hadamard product (element-wise multiplication).

CNN-BiLSTM, on the other hand, is a hybrid model that combines the capabilities of Convolutional Neural Networks (CNN) with the bidirectional variant of LSTM (BiLSTM). While LSTM is primarily designed to capture long-term dependencies in sequential data, BiLSTM extends this by processing data in both forward and backward directions. This bidirectional approach allows the network to consider both past and future information, making it more effective for certain time series tasks.

CNNs are particularly used at extracting spatial features from data by applying multiple filters over the input. These filters capture essential patterns in the data through the convolutional and pooling layers. As the data passes through successive CNN layers, increasingly abstract features are obtained, which can then be passed on to the BiLSTM for further temporal analysis. The CNN component extracts localized features, while the BiLSTM processes these features across time steps, capturing long-range dependencies in both directions.

In the CNN-BiLSTM model, the CNN layers first extract spatial features from the input data, which are then fed into the BiLSTM to capture temporal dependencies. The hidden states from both the forward and backward LSTM layers are combined to form a comprehensive feature representation:

$$\overrightarrow{h}_t = LSTM\left(\overrightarrow{h}_{t-1}, x_t, c_{t-1}\right) \tag{22}$$

$$\overleftarrow{h}_t = LSTM\left(\overleftarrow{h}_{t+1}, x_t, c_{t+1}\right) \tag{23}$$

The final hidden state is a concatenation of both directional outputs, given by $H_t = [\overrightarrow{h}_t, \overleftarrow{h}_t]$. This hybrid architecture allows the model to represent both spatial and temporal patterns in the data, making it particularly effective for financial time series forecasting, where both types of dependencies are critical for accurate predictions.

4.3 Model Confidence Set (MCS) Test

While the various evaluation metrics can rank forecasts, they do not easily indicate if the differences in values are statistically significant. To address this issue, we utilize the Model Confidence Set (MCS) method introduced by Hansen et al. (xxxx). The MCS method involves a sequence of statistical tests to identify a set of 'superior' models. By sequentially excluding the least effective model based on the Equal Predictive Ability (EPA) criterion, the MCS method narrows down to the optimal model(s) within a specified confidence level.

The MCS procedure begins with the initial set of candidate forecasting models, denoted as $M_0 = \{1, 2, \dots, m_0\}$. To compare the performance of these models, the loss differential $d_{ij,t}$ between models i and j at time t is calculated as:

$$d_{ij,t} = L(\sigma^2_{RV,t}, \sigma^2_{i,t}) - L(\sigma^2_{RV,t}, \sigma^2_{j,t}) \tag{24}$$

where L represents the chosen loss function, $\sigma^2_{RV,t}$ denotes the realized volatility, and $\sigma^2_{i,t}$, $\sigma^2_{j,t}$ are the forecasted volatilities by models i and j, respectively.

The null hypothesis H_0 of Equal Predictive Ability (EPA) is formulated as:

$$H_0 : \mathbb{E}[d_{ij,t}] = 0, \text{for all i,j} \in M_s \tag{25}$$

where $M_s \subseteq M$ represents the current set of models under consideration. At each stage, H_0 is tested using a predetermined significance level α. If H_0 is rejected, indicating significant differences in predictive accuracy, the model contributing most to the test statistic is eliminated from M_s, and the procedure is repeated until no further models can be eliminated. If the null hypothesis is not rejected, the set of models $\widehat{M}_{best}$ is considered to be M_s. The resulting $\widehat{M}_{best}$ is referred to as the Model Confidence Set (MCS). This set of best models is determined based on a user-defined criterion for the loss function. In our analysis, we utilize Mean Squared Error (MSE) as well as quasi-likelihood (QLIKE) as the criteria.

4.4 Forecast Comparison Test (DM Test)

To validate our empirical results and comprehensively evaluate forecast quality, we utilize the Diebold-Mariano (DM) test, proposed by Diebold and Mariano (2002). Traditional accuracy measures, such as Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE), can lead to incorrect conclusions when comparing model performance if statistical significance tests are not conducted. To address this issue, Diebold and Mariano (2002) introduced a test to statistically validate the performance of a model against a benchmark. The DM test examines the null hypothesis of no difference in forecast accuracy between two competing models and can be applied using various error criteria, including straight differences, absolute differences, or squared differences.

Let the forecast errors for models A and B be represented by $e_i^{(A)} = y_i - \hat{y}_i^{(A)}$ and $e_i^{(B)} = y_i - \hat{y}_i^{(B)}$ for $i \in \{1, 2, \dots N\}$. The loss function applied to the forecast errors is defined as $L(e_i^{(A)}) = (e_i^{(A)})^2$ and similarly for $L(e_i^{(A)})$. The Diebold-Mariano (DM) test statistic is computed using

$$DM = \frac{\sum_{i=1^N} \left(L(e_i^{(A)}) - L(e_i^{(B)}) \right) / N}{\sqrt{\widehat{S}^2 / N}} \tag{26}$$

where $\widehat{S}^2$ is the estimated variance of the difference in losses. The expected difference in loss functions is denoted as $\mathbb{E}[d] = L(e_i^{(A)}) - L(e_i^{(B)})$.

The hypotheses for the Diebold-Mariano test are:

- H0: The loss functions of the predictions are not statistically different ($\mathbb{E}[d] = 0$)
- H1: Model A has better predictive performance than model B ($\mathbb{E}[d] > 0$)
- H2: Model B has better predictive performance than model A ($\mathbb{E}[d] < 0$)

5 Results

5.1 Baseline Results

We begin our analysis by examining the sentiment data as a predictor of subsequent realized volatility. To do this, we conduct a regression analysis using Eq. 2, utilizing the entire dataset for each cryptocurrency respectively. On average, variance inflation factors of the regressors are less than 2, multicollinearity therefore is not an issue. The results of this analysis are presented in Table 3. The findings reveal that various aspects of sentiment data, including the number, scores, and density of general crypto news, exhibit a significant association with subsequent realized volatility across nearly all coins in the sample. Furthermore, the number of news on corona and crypto regulation also demonstrates predictive power. Notably, the inclusion of sentiment data greatly enhances the model fit, as indicated by the improved adjusted R-squared value, in comparison to a model that excludes sentiment data. Hence, sentiment seems a worthwhile candidate for predicting volatility.

Next we turn to one-step ahead forecasts of realized volatility. The accuracy of forecasts is being assessed by loss functions, i.e., mean absolute error (MAE), root mean squared error (RMSE), as well as quasi-likelihood (QLIKE). The results, presented in Table 4, unveil key insights.

Firstly, when sentiment data is not considered (upper panel of Table 4), we find that the benchmark HAR model performs well only for BTC. For other cryptocurrencies we find improved accuracy when utilizing machine learning (ML) models, particularly the CNN-BiLSTM model. Generally, gray shaded cells in Table 4 highlight the best performing model for each coin and loss function respectively. Secondly, when sentiment data is incorporated (lower panel of Table 4), the superiority of ML models is further confirmed, with LightGBM and XGBoost emerging as the top-performing models in terms of forecast accuracy. Overall, we assess forecast accuracy for three loss functions, six models, and eight coins, a total of 144 evaluated cases. In 78 of these cases or 54.17% (cells in the lower panel tagged with an asterisk), adding sentiment improves the accuracy of the respective model compared to models without sentiment.

Interestingly, and in contrast to the findings of Sapkota (2022), our findings reveal that the inclusion of sentiment data does not increase the accuracy of the HAR model with bitcoin being the only exception. Furthermore, while the CNN-BiLSTM model performs well when excluding sentiment, the inclusion of sentiment data, on average across the eight coins under consideration, markedly deteriorates the loss function values. The same is true for the MLP model. Conversely, sentiment on average significantly improves forecast accuracy in terms of MAE, RMSE as well

Table 3 OLS full sample results. This table reports full sample results for the regression specified in Eq. 2. Panels show estimated coefficients of lagged realized volatilities (lagged RV), the total number of news, (Number), the average sentiment scores (Scores), as well as the density of news (Density). Number, Scores and Density include 6 sentiment time series each, i.e., general crypto new, crpyto mining, corona, news on regulation, influencer, as well as coin specific news (i.e., bitcoin news for BTC, and so forth, denoted by [Coin]). The bottom panel reports R^2 and adjusted R^2 for the full model including sentiment data (w/ S), as well as the corresponding R^2 and adjusted R^2 when sentiment data is discarded (no S). ***/**/* denote statistical significance at the 1/5/10 percent level

	Variable	BTC	ETH	ADA	DOGE	DOT	LTC	SHIB	SOL
lagged RV	constant	−0.0047**	−0.0048**	−0.0041*	−0.0044	−0.0030	−0.0023	0.0021	−0.007**
	RV_{t-1}	0.3387***	0.3287***	0.4629***	0.3750***	0.3584***	0.3346***	0.384***	0.4983***
	$RV^{(8)}_{t-1)}$	0.4120***	0.4787***	0.371***	0.4136***	0.5062***	0.4017***	0.4420***	0.3537***
	$RV^{(28)}_{t-1}$	−0.0010	−0.0468	−0.0281	−0.1048**	−0.0502	−0.0310	−0.0401	−0.1054***
Number	General	0.0047***	0.0087***	0.0062***	0.0014	0.0086***	0.0053***	0.0038	0.0075***
	Mining	−0.0021	0.0076	−0.0035	−0.0325	−0.0118	−0.0075	−0.0282	−0.058*
	Corona	−0.0008***	−0.001***	−0.0003	−0.0008**	−0.0003	−0.0004*	−0.0008**	−0.0009***
	Regulation	−0.0201*	−0.0207	−0.0281**	−0.0181	−0.0351**	−0.0511***	−0.0306	−0.0304
	Influencer	0.0194***	0.0243***	0.0365***	0.0472***	0.0340***	0.0371***	0.0384***	0.0545***
	[Coin]	0.0107**	−0.0024	−0.0336	0.1412***	−0.0825**	0.0074	0.0656**	0.0502**
Scores	General	−0.0058	−0.0157**	−0.0167**	−0.0409***	−0.0131*	−0.0372***	−0.0226**	−0.0475***
	Mining	−0.0005	−0.0006	−0.0009	−0.0010	−0.0009	−0.0014	−0.0014	−0.0006
	Corona	−0.0149	−0.0108	0.0088	0.0199	0.0138	−0.0109	0.0223	0.0010
	Regulation	0.0025*	0.0025	0.0021	0.0045*	0.0019	0.0031	0.0024	0.0054**
	Influencer	−0.0016	−0.0030	−0.0030	0.0013	−0.0015	−0.0051	0.0019	0.0009
	[Coin]	−0.0034	−0.0037	0.0012	0.0011	−0.0021	0.0018	−0.0019	0.0004
Density	General	0.0255***	0.0365***	0.0262***	0.0408***	0.0237**	0.0408***	0.0134	0.0616***
	Mining	0.0007	0.0007	−0.0008	−0.0035*	−0.0012	−0.0009	−0.0015	−0.0005
	Corona	−0.0200**	−0.0303***	−0.0147	0.0051	−0.0178	−0.0184	−0.0218	−0.0210
	Regulation	0.0029*	0.0013	0.0017	0.0015	−0.0009	0.0019	0.0021	0.0025
	Influencer	−0.0002	0.0002	0.0055	0.0034	0.0010	0.0027	0.0007	−0.0002

Table 3 (continued)

	Variable	BTC	ETH	ADA	DOGE	DOT	LTC	SHIB	SOL
	[Coin]	−0.0030	0.0003	−0.0018	0.0046	0.0012	0.0006	0.0080***	−0.0031
w/ S	R^2	0.5013	0.5320	0.6073	0.5315	0.5888	0.4771	0.5019	0.6481
	adj. R^2	0.4941	0.5252	0.6016	0.5247	0.5828	0.4695	0.4947	0.6429
no S	R^2	0.4172	0.4598	0.5715	0.4886	0.5472	0.4256	0.4756	0.6029
	adj. R^2	0.4160	0.4587	0.5706	0.4876	0.5462	0.4244	0.4745	0.6021
obs		1,461	1,461	1,461	1,461	1,461	1,461	1,461	1,461

Table 4 Forecasting results. This table reports results from the one-step ahead forecasts for realized volatility. The upper (lower) panel shows values for mean absolute error (MAE), root mean squared error (RMSE) and quasi-likelihood (QLIKE) for the benchmark HAR model as well as the 5 machine learning models excluding (including) sentiment data (Sent. - no/yes). Columns depict data for each of the 8 coins in the sample. Gray shaded cells mark the lowest value for each loss function per coin, i.e., 0.0047 MAE for BTC excluding sentiment points at the HAR model being the most accurate model among the six candidates. Cells in the lower panel marked with a * indicate improved forecast accuracy when sentiment is added to the same model and coin

Loss f	Model	Sent	BTC	ETH	ADA	DOGE	DOT	LTC	SHIB	SOL
MAE	HAR	no	*0.0047*	0.0075	0.0087	0.0188	0.0077	0.0113	0.0188	0.0175
	LightGBM		0.0070	0.0073	0.0084	*0.0101*	0.0080	*0.0070*	0.0090	0.0096
	XGBoost		0.0069	0.0089	0.0063	0.0106	0.0097	0.0085	0.0108	0.0113
	LSTM		0.0077	0.0087	0.0090	0.0136	0.0136	0.0094	0.0127	0.0138
	MLP		0.0091	0.0061	0.0063	0.0153	*0.0061*	0.0095	0.0104	0.0104
	CNN-BiLSTM		0.0054	*0.0060*	*0.0054*	0.0120	0.0067	0.0072	*0.0080*	*0.0099*
RMSE	HAR	no	*0.0070*	0.0093	0.0105	0.0251	0.0093	0.0135	0.0221	0.0264
	LightGBM		0.0101	0.0121	0.0142	0.0170	0.0160	0.0125	0.0166	0.0185
	XGBoost		0.0076	0.0093	0.0092	*0.0121*	0.0102	*0.0092*	0.0129	*0.0127*
	LSTM		0.0095	0.0110	0.0110	0.0201	0.0150	0.0122	0.0168	0.0193
	MLP		0.0102	0.0089	0.0096	0.0211	0.0092	0.0132	0.0155	0.0171
	CNN-BiLSTM		0.0074	*0.0089*	*0.0083*	0.0178	*0.0086*	0.0110	*0.0117*	0.0167
QLIKE	HAR	no	−7.3555	−4.8326	−3.4134	9.8868	−4.1559	−0.9988	9.0041	2.1850
	LightGBM		−7.1672	−6.6863	−5.9918	−6.1298	−6.0037	−6.4401	−5.5796	−5.7645
	XGBoost		−7.1158	−6.6758	−6.5879	*−6.2481*	−5.9363	−6.3937	−5.5694	−5.7678
	LSTM		−7.7244	−7.0343	−6.7812	−5.9957	−6.5700	−6.4974	−6.2544	−6.0272
	MLP		−7.5091	−7.3045	−7.1396	−6.0163	*−7.1102*	−6.6232	−6.2204	−6.2390
	CNN-BiLSTM		*−7.9721*	*−7.3277*	*−7.1882*	−6.0497	−7.0910	*−6.6981*	*−6.4366*	*−6.3322*
Loss f	Model	Sent	BTC	ETH	ADA	DOGE	DOT	LTC	SHIB	SOL
MAE	HAR	yes	*0.0046**	0.0082	0.0093	0.0201	0.0081	0.0119	0.0195	0.0188
	LightGBM		0.0050*	0.0059*	*0.0058**	*0.0092**	*0.0053**	*0.0064**	*0.0092*	0.0096*

Table 4 (continued)

Loss f	Model	Sent	BTC	ETH	ADA	DOGE	DOT	LTC	SHIB	SOL
	XGBoost		0.0053*	0.0065*	0.0062*	0.0106*	0.0063*	0.0071*	0.0109	*0.0096**
	LSTM		0.0049*	*0.0056**	0.0092	0.0116*	0.0134*	0.0071*	0.0093*	0.0134*
	MLP		0.0058*	0.0086	0.0084	0.0194	0.0075	0.0099	0.0098*	0.0303
	CNN-BiLSTM		0.0062	0.0061	0.0082	0.0105*	0.0137	0.0064*	0.0104	0.0106
RMSE	HAR	yes	*0.0066**	0.0100	0.0111	0.0266	0.0098	0.0141	0.0228	0.0274
	LightGBM		0.0069*	0.0085*	*0.0084**	0.0153*	*0.0074**	0.0096*	*0.0119**	0.0189
	XGBoost		0.0072*	0.0085*	0.0089*	*0.0122*	0.0096*	*0.0086**	0.0128*	*0.0116**
	LSTM		0.0069*	*0.0084**	0.0111	0.0179*	0.0150	0.0105*	0.0126*	0.0235
	MLP		0.0093	0.0138	0.0132	0.0271	0.0115	0.0144	0.0133	0.0389
	CNN-BiLSTM		0.0079	0.0091	0.0102	0.0158*	0.0154	0.0094*	0.0128	0.0200
QLIKE	HAR	yes	−7.2277	−4.3143	−3.1234	11.4014	−4.0381	−0.4069	9.2322	3.1213
	LightGBM		−8.0206*	−7.3199*	−7.1087*	*−6.3138**	−7.2515	−6.7399*	*−6.3573**	−6.3948*
	XGBoost		−7.2107*	−7.2872*	*−7.1123**	−6.1178	*−7.2815**	−6.7452*	−6.1951*	*−6.3967**
	LSTM		−8.0761*	*−7.3873**	−6.7800	−6.1600*	−6.6075*	−6.7010*	−6.3093*	−6.2731*
	MLP		*−8.0838**	−7.2002	−7.0114	−5.7917	−7.1885*	−6.5903	−6.3307*	−5.4129
	CNN-BiLSTM		−7.8547	−7.3045	−6.8661	−6.0919*	−6.6103	*−6.7757**	−6.1967	−6.3487*

as QLIKE with LightGBM, XGBoost and LSTM. Table 5 reports detailed results of average changes of the loss function values across the eight coins in the sample.

In summary, when it comes to predicting realized volatility of cryptocurrencies, the final recommendation therefore tends towards using either of the latter three machine learning tools (LightGBM, XGBoost, LSTM) along with the inclusion of sentiment data.

5.2 Comparing models' forecasting ability

Table 6 shows the MCS test results using both MSE and QLIKE metrics derived from out-of-sample forecasting. Bold values in the table indicate the optimal model selected by MCS. The test also identifies other models with Equal Predictive Ability (EPA) at the specified confidence level. Our findings indicate that before the inclusion of sentiment data (panels 1 and 3 of the table), the LightGBM model is consistently identified as one of the superior models for ETH, DOGE, and LTC demonstrating lower prediction errors and higher stability in forecasts. The XGBoost model also shows strong performance across several cryptocurrencies. However, the HAR model, while performing well for BTC and DOGE, generally underperforms compared to machine learning models for other cryptocurrencies.

After incorporating sentiment data (panels 2 and 4 of the table), the performance of machine learning models, particularly XGBoost, further improves. The XGBoost model shows significant improvements on five occasions, while the HAR model shows notable improvement in its MCS ranking for BTC and DOGE with sentiment data but still lags behind the advanced machine learning models for most other cryptocurrencies. This outcome underscores the efficacy of machine learning and gradient boosting techniques in capturing the complex dynamics of cryptocurrency volatility, especially when sentiment data is included.

Moreover, the MCS test results for individual forecasts using the QLIKE loss function, both without sentiment and with sentiment, show that ML models, particularly LSTM and CNNBiLSTM, consistently outperform the other models. Including sentiment data further supports the superiority of ML models across all cryptocurrencies, while the HAR model is eliminated in all cases. These findings underscore

Table 5 **Average relative loss function values after sentiment inclusion**. This table reports the ratio of average loss function values across all coins with and without the inclusion of sentiment. For instance, the value 1.0533 for the HAR model and MAE implies that the inclusion of sentiment on average increases MAE by 5.33% across all coins

	MAE	RMSE	QLIKE
HAR	1.0533	1.0378	1.1571
LightGBM	0.8374	0.7306	0.8979
XGBoost	0.8519	0.9533	0.9268
LSTM	0.8236	0.9025	0.9748
MLP	1.3461	1.3234	1.0162
CNN-BiLSTM	1.2354	1.1417	1.0194

Table 6 MCS test results. This table presents the results of the MCS test conducted using both MSE and QLIKE loss functions. The confidence level for the test is set at $\alpha = 0.2$ Models demonstrating Equal Predictive Ability (EPA) at the 75% level are bolded in the table

	BTC	ETH	ADA	DOGE	DOT	LTC	SHIBA	SOL
MSE (without sentiment)								
HAR	0.0000	0.0000	0.0000	0.6636	0.0000	0.2950	0.2940	0.4700
LightGBM	0.0000	**1.0000**	0.0000	**1.0000**	0.2626	**1.0000**	0.9702	0.2954
XGBoost	**1.0000**	0.5432	**1.0000**	0.7686	0.7432	0.4456	0.1616	**1.0000**
LSTM	0.7182	0.7878	0.5076	0.6218	0.9274	0.7120	0.9630	0.6486
CNNBiLSTM	**1.0000**	0.0566	**1.0000**	0.2408	0.5536	0.0852	0.0034	0.5680
MLP	0.9868	0.7630	0.9876	0.6766	**1.0000**	0.7042	**1.0000**	0.7726
MSE (with sentiment)								
HAR	0.4922	0.0236	0.0178	**1.0000**	0.0394	0.8212	0.0638	**1.0000**
LightGBM	0.7724	0.2540	0.0000	0.4368	0.8442	**1.0000**	0.3370	0.7974
XGBoost	**1.0000**	**1.0000**	**1.0000**	0.8768	**1.0000**	0.7948	**1.0000**	0.9602
LSTM	0.7408	0.2774	0.3976	0.2356	0.3122	0.3692	0.5154	0.3176
CNNBiLSTM	**1.0000**	0.2758	0.3244	0.0000	0.8442	0.0306	0.1018	0.0482
MLP	0.0000	0.7398	0.8076	0.8380	0.8600	0.4712	0.9928	0.7588
	BTC	ETH	ADA	DOGE	DOT	LTC	SHIBA	SOL
QLIKE (without sentiment)								
HAR	0.0000	0.0000	0.0000	0.7770	0.0000	0.5802	0.0000	0.0000
LightGBM	0.0000	0.0000	0.0000	0.0612	0.3960	0.1668	0.0000	0.0000
XGBoost	0.0000	0.0000	0.0000	0.0368	0.0000	0.0098	0.0000	0.0000
LSTM	0.0000	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
CNNBiLSTM	**1.0000**	0.0062	0.9534	0.1904	0.0523	0.0140	0.0000	0.1738
MLP	0.0000	0.0630	0.0000	0.9458	0.0047	0.8162	0.0000	0.5946
QLIKE (with sentiment)								
HAR	0.0000	0.0000	0.0000	0.5588	0.0000	0.0960	0.0000	0.0904
LightGBM	0.6608	0.0000	0.0000	0.7096	**1.0000**	**1.0000**	0.5056	**1.0000**
XGBoost	0.0032	**1.0000**	**1.0000**	0.9254	0.9952	0.5040	**1.0000**	0.8218
LSTM	0.9786	0.0000	0.5392	0.2456	0.9186	0.3856	0.3562	0.3648
CNNBiLSTM	**1.0000**	0.0000	0.1268	0.2232	0.9952	0.0138	0.0000	0.0506
MLP	0.0164	0.6394	0.9600	**1.0000**	0.4256	0.3480	0.9656	0.9874

the superior forecasting accuracy of advanced ML models, highlighting their ability to leverage complex patterns and diverse data sources for improved financial predictions.

Tables 7 and 8 further provide a comparative analysis of the forecast accuracy in terms of MSE and QLIKE between different model pairs. The DM test assesses the null hypothesis of equal predictive accuracy, thereby offering insights into whether observed differences in forecast performance are statistically significant.

Contrary to the MCS test, results derived from the DM test are ambiguous. When evaluating forecast accuracy in terms of MSE, the HAR model performs reasonably well and is superior to LBM and XGBoost when discarding and including sentiment

Table 7 Diebold-Mariano Test Results. This table reports results from the pairwise Diebold-Mariano test, where one model's forecasting accuracy in terms of mean-squared error (MSE) is related to another model from our set of candidate models

Diebold-Mariano MSE (without sentiment)

		BTC		ETH		ADA		DOGE		DOT		LTC		SHIB		SOL	
Model A	Model B	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val
HAR	LBM	1.357	0.176	1.239	0.217	−1.200	0.231	1.290	0.198	1.197	0.232	1.223	0.222	0.896	0.371	1.079	0.282
HAR	XGB	1.632	0.104	1.208	0.228	1.196	0.233	1.318	0.189	1.197	0.232	1.116	0.265	0.842	0.400	1.257	0.210
HAR	LSTM	2.026	0.044	0.428	0.669	0.599	0.550	−0.343	0.732	0.457	0.648	−0.518	0.605	−0.244	0.807	−1.213	0.226
HAR	CNNBiLSTM	1.967	0.050	−0.314	0.754	1.537	0.125	−0.902	0.368	1.124	0.262	−1.007	0.315	−1.134	0.258	−1.222	0.223
HAR	MLP	1.692	0.092	1.273	0.204	1.202	0.230	1.286	0.199	1.229	0.220	1.138	0.256	1.009	0.314	1.396	0.164
LBM	XGB	1.764	0.079	−1.340	0.181	1.349	0.178	−1.188	0.236	−1.007	0.318	−1.732	0.084	−1.669	0.096	1.356	0.176
LBM	LSTM	1.313	0.190	−1.112	0.267	1.917	0.056	−1.343	0.180	−0.897	0.370	−1.282	0.201	−1.068	0.287	−1.375	0.170
LBM	CNNBiLSTM	2.215	0.028	−1.277	0.203	1.674	0.095	−1.368	0.172	−0.791	0.430	−1.381	0.168	−1.188	0.236	−1.413	0.159
LBM	MLP	1.896	0.059	−1.107	0.269	1.394	0.164	−1.288	0.199	0.513	0.608	−1.249	0.213	−0.403	0.687	1.419	0.157
XGB	LSTM	−1.034	0.302	−1.043	0.298	−1.034	0.302	−1.359	0.175	−0.897	0.370	−1.210	0.227	−1.010	0.314	−1.372	0.171
XGB	CNNBiLSTM	−0.765	0.445	−1.251	0.212	−0.765	0.445	−1.379	0.169	−0.791	0.430	−1.330	0.185	−1.144	0.254	−1.397	0.164
XGB	MLP	−1.044	0.297	−0.121	0.904	−1.044	0.297	−1.122	0.263	0.513	0.608	−1.069	0.286	−0.265	0.791	−0.649	0.517
LSTM	CNNBiLSTM	1.344	0.180	−1.529	0.127	1.344	0.180	−1.359	0.175	0.993	0.321	−2.110	0.036	−1.485	0.139	−1.131	0.259
LSTM	MLP	1.030	0.304	1.114	0.266	1.030	0.304	1.363	0.174	0.989	0.324	1.255	0.211	1.350	0.178	1.443	0.150
CNNBiLSTM	MLP	0.616	0.539	1.336	0.183	0.616	0.539	1.388	0.166	0.876	0.382	1.439	0.151	1.485	0.139	1.460	0.145

Diebold-Mariano MSE (with sentiment)

		BTC		ETH		ADA		DOGE		DOT		LTC		SHIB		SOL	
Model A	Model B	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val
HAR	LBM	1.295	0.196	0.748	0.455	−1.657	0.099	0.275	0.784	0.287	0.775	1.048	0.295	1.071	0.285	−0.429	0.668
HAR	XGB	1.264	0.207	1.271	0.205	1.019	0.309	0.509	0.612	0.817	0.415	−0.012	0.991	1.117	0.265	−0.604	0.546

Table 7 (continued)

Diebold-Mariano MSE (with sentiment)

		BTC		ETH		ADA		DOGE		DOT		LTC		SHIB		SOL	
HAR	LSTM	0.636	0.525	−1.389	0.166	−1.355	0.176	−1.649	0.100	−1.275	0.204	−1.630	0.104	−1.500	0.135	−1.589	0.113
HAR	CNNBiLSTM	−0.970	0.333	−1.391	0.165	−1.320	0.188	−1.719	0.087	−1.234	0.218	−1.638	0.102	−1.517	0.131	−1.624	0.106
HAR	MLP	0.317	0.752	0.548	0.584	−0.403	0.687	0.332	0.740	−0.210	0.834	−0.529	0.597	0.720	0.472	−0.471	0.638
LBM	XGB	1.122	0.263	1.563	0.119	1.719	0.087	0.412	0.681	1.025	0.306	−1.416	0.158	1.150	0.251	0.013	0.990
LBM	LSTM	−1.221	0.223	−1.786	0.075	1.953	0.052	−1.724	0.086	−1.228	0.220	−1.805	0.072	−1.627	0.105	−1.773	0.077
LBM	CNNBiLSTM	−1.251	0.212	−1.812	0.071	2.163	0.031	−1.686	0.093	−1.173	0.242	−1.809	0.072	−1.583	0.115	−1.763	0.079
LBM	MLP	−1.625	0.105	0.032	0.975	1.960	0.051	0.394	0.694	−0.558	0.577	−1.910	0.057	0.234	0.815	−0.420	0.675
XGB	LSTM	−1.719	0.087	−1.759	0.080	−1.605	0.110	−1.665	0.097	−1.415	0.158	−1.806	0.072	−1.578	0.116	−1.715	0.087
XGB	CNNBiLSTM	−1.239	0.216	−1.797	0.073	−1.579	0.115	−1.671	0.096	−1.333	0.183	−1.809	0.071	−1.540	0.125	−1.723	0.086
XGB	MLP	−1.879	0.061	−0.743	0.458	−0.956	0.340	−0.161	0.872	−0.924	0.357	−1.109	0.268	−0.066	0.948	−0.269	0.788
LSTM	CNNBiLSTM	−0.824	0.410	−1.036	0.301	0.591	0.555	−1.102	0.271	−0.851	0.395	−1.749	0.081	−1.117	0.265	−0.921	0.358
LSTM	MLP	−1.841	0.067	1.769	0.078	1.504	0.134	1.727	0.085	1.697	0.091	1.745	0.082	1.440	0.151	1.688	0.092
CNNBiLSTM	MLP	0.620	0.536	1.766	0.079	1.586	0.114	1.676	0.095	1.535	0.126	1.752	0.081	1.559	0.120	1.679	0.094

Table 8 Diebold-Mariano Test Results. This table reports results from the pairwise Diebold-Mariano test, where one model's forecasting accuracy in terms of quasi likelihood (QLIKE) is related to another model from our set of candidate models

Diebold-Mariano QLIKE (without sentiment)

		BTC		ETH		ADA		DOGE		DOT		LTC		SHIB		SOL	
Model A	Model B	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val
HAR	LBM	2.036	0.042	−0.760	0.448	−1.969	0.049	−1.040	0.298	−1.795	0.073	−0.503	0.615	−2.528	0.011	−1.105	0.269
HAR	XGB	1.826	0.068	−1.185	0.236	−1.432	0.152	−1.522	0.128	−1.795	0.073	−1.058	0.290	−2.754	0.006	−1.122	0.262
HAR	LSTM	3.919	0.000	3.567	0.000	3.279	0.001	1.209	0.227	3.470	0.001	1.636	0.102	3.414	0.001	1.740	0.082
HAR	CNNBiLSTM	4.101	0.000	1.724	0.085	3.591	0.000	0.291	0.771	3.745	0.000	0.133	0.894	−0.408	0.683	1.451	0.147
HAR	MLP	3.367	0.001	1.182	0.237	0.683	0.494	0.448	0.654	2.313	0.021	0.902	0.367	1.411	0.158	1.374	0.170
LBM	XGB	−0.435	0.663	−2.706	0.007	0.619	0.536	−2.844	0.004	3.502	0.000	−3.226	0.001	−2.511	0.012	−0.251	0.802
LBM	LSTM	3.477	0.001	3.288	0.001	5.969	0.000	1.693	0.090	3.977	0.000	1.800	0.072	3.192	0.001	2.152	0.031
LBM	CNNBiLSTM	3.985	0.000	2.061	0.039	4.773	0.000	0.888	0.374	4.262	0.000	0.432	0.666	1.631	0.103	1.878	0.060
LBM	MLP	3.295	0.001	4.645	0.000	2.264	0.024	2.943	0.003	4.368	0.000	2.257	0.024	3.381	0.001	3.074	0.002
XGB	LSTM	3.407	0.001	3.502	0.000	3.901	0.000	1.935	0.053	3.977	0.000	2.077	0.038	3.263	0.001	2.095	0.036
XGB	CNNBiLSTM	3.895	0.000	2.383	0.017	4.814	0.000	1.161	0.246	4.262	0.000	0.758	0.448	1.888	0.059	1.829	0.067
XGB	MLP	3.269	0.001	4.703	0.000	5.958	0.000	3.267	0.001	4.368	0.000	2.787	0.005	3.455	0.001	3.139	0.002
LSTM	CNNBiLSTM	1.315	0.188	−4.207	0.000	−0.072	0.943	−2.104	0.035	−0.464	0.643	−4.034	0.000	−5.335	0.000	−1.758	0.079
LSTM	MLP	−3.065	0.002	−2.297	0.022	−2.756	0.006	−0.951	0.342	−2.821	0.005	−1.279	0.201	−2.226	0.026	−1.064	0.287
CNNBiLSTM	MLP	−3.998	0.000	−0.380	0.704	−3.584	0.000	0.025	0.980	−3.267	0.001	0.520	0.603	1.182	0.237	−0.731	0.465

Diebold-Mariano QLIKE (with sentiment)

		BTC		ETH		ADA		DOGE		DOT		LTC		SHIB		SOL	
Model A	Model B	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val	DM	p-val
HAR	LBM	2.738	0.006	4.263	0.000	−2.959	0.003	1.277	0.202	3.791	0.000	2.986	0.003	4.574	0.000	2.966	0.003
HAR	XGB	1.791	0.073	4.647	0.000	4.355	0.000	1.468	0.142	4.317	0.000	2.534	0.011	4.313	0.000	−0.604	0.006

Table 8 (continued)

Diebold-Mariano QLIKE (with sentiment)

		BTC		ETH		ADA		DOGE		DOT		LTC		SHIB		SOL	
HAR	LSTM	3.275	0.001	0.656	0.512	1.496	0.135	−1.447	0.148	2.108	0.035	−0.679	0.497	1.455	0.146	−0.431	0.666
HAR	CNNBiLSTM	3.329	0.001	−0.327	0.744	1.255	0.210	−1.701	0.089	1.814	0.070	−1.071	0.284	0.050	0.960	−0.701	0.483
HAR	MLP	1.682	0.093	3.575	0.000	2.865	0.004	1.592	0.111	1.682	0.093	2.342	0.019	3.401	0.001	2.261	0.024
LBM	XGB	−2.450	0.014	2.170	0.030	5.702	0.000	0.133	0.894	−0.294	0.769	−1.565	0.118	0.956	0.339	0.133	0.010
LBM	LSTM	1.167	0.243	−2.555	0.011	6.149	0.000	−2.594	0.009	−1.168	0.243	−2.775	0.006	−2.167	0.030	−2.599	0.009
LBM	CNNBiLSTM	1.099	0.272	−2.649	0.008	6.219	0.000	−2.555	0.011	−0.910	0.363	−3.110	0.002	−2.828	0.005	−2.765	0.006
LBM	MLP	−0.535	0.592	0.390	0.697	6.042	0.000	1.476	0.140	−1.627	0.105	−2.448	0.014	0.045	0.964	−0.575	0.566
XGB	LSTM	1.854	0.064	−2.658	0.008	−1.737	0.082	−2.426	0.015	−0.871	0.384	−2.721	0.007	−2.259	0.024	−1.715	0.000
XGB	CNNBiLSTM	1.551	0.121	−2.727	0.006	−2.080	0.038	−2.466	0.014	−0.705	0.481	−3.087	0.002	−2.928	0.003	−1.723	0.006
XGB	MLP	0.301	0.763	−0.568	0.570	−0.413	0.680	0.952	0.341	1.440	0.151	−0.986	0.324	−0.280	0.779	−0.269	0.008
LSTM	CNNBiLSTM	0.488	0.626	−1.974	0.048	−1.862	0.063	−1.325	0.185	−0.058	0.954	−4.134	0.000	−2.461	0.014	−1.882	0.060
LSTM	MLP	−2.581	0.010	2.964	0.003	1.930	0.054	3.233	0.001	1.697	0.091	2.668	0.008	2.173	0.030	2.661	0.008
CNNBiLSTM	MLP	−1.379	0.168	2.899	0.004	2.485	0.013	3.004	0.003	1.535	0.126	3.060	0.002	3.077	0.002	2.809	0.005

respectively (as evidenced by mainly positive values for the DM test statistic). The other ML models provide more accurate forecasts as compared to the HAR model. In other words, in 28 out of 48 comparisons of the HAR model with either of LSTM, CNNBiLSTM, and MLP, the machine learning approach is more reliable in terms of forecast accuracy. However, only a few of the results presented in Table 7 are statistically significant. Interestingly though, when referring to QLIKE as the pivotal measure of forecast accuracy, the HAR model significantly outperforms all of the ML models for bitcoin when sentiment is discarded. For the remaining coins, forecasts without the inclusion of sentiment are consistently more accurate when LBM or XGBoost are being applied. The remaining ML perform poorly and the HAR model often provides statistically significant increased precision. Upon the inclusion of sentiment, forecast precision as measured by QLIKE still points at the HAR model as a feasible tool. However, in some few instances, ML models provide more accurate forecasts.

In conclusion, the analysis of MCS and DM tests largely confirms our earlier forecasting results, highlighting the advantage of advanced machine learning models in volatility forecasting and underscoring the pivotal role of sentiment data in enhancing predictive accuracy.

6 Conclusion

In this study, we investigated the distinct role of news sentiment in enhancing the accuracy of cryptocurrency volatility forecasts. By collecting crypto-specific news headlines, our research seeks to understand how sentiment embedded in news can inform and improve volatility predictions. We employ the Heterogeneous Autoregressive (HAR) model as a benchmark for forecasting the 6-hour realized volatility across eight major cryptocurrencies. We then compare against five advanced machine learning (ML) models to assess their relative performance.

The out-of-sample findings show that the machine learning models consistently outperformed the HAR benchmark across the board, with the notable exception of Bitcoin. This exception presents an intriguing case for further investigation, considering Bitcoin's dominance, unique market dynamics, and investor perception. Moreover, the integration of sentiment data as an explanatory variable resulted in a substantial improvement in forecasting accuracy in over half (54%) of the cases evaluated. Specifically, this improvement shows the effectiveness of sentiment analysis with ML in capturing market dynamics that linear models such as the HAR may overlook. When assessing forecast accuracy using established tools like the MCS test, the baseline results can be validated. Machine learning models that have already been shown to outperform the HAR model continue to deliver more accurate predictions of realized volatility. However, the DM test provides a nuanced perspective on these findings. For both loss functions-MSE and QLIKE-as well as forecasts with and without sentiment data, the HAR model can still generate reasonably accurate forecasts. While some machine learning approaches do outperform the benchmark HAR model, there is no definitive choice among the respective ML models. In recent years, the DM test has become a standard for comparing forecasts in out-of-sample environments, but its primary purpose is not to compare different models such as HAR and ML models (Diebold (2015)). Therefore, the main findings of this study should be viewed within the context of the various analyses conducted. Overall, incorporating sentiment data and utilizing machine learning tools for forecasting realized volatility in cryptocurrency markets proves to be beneficial.

The findings of this study align with existing research, illustrating that investor sentiment significantly influences the cryptocurrency market. This is in line with prior studies by Jo et al. (2020), Conghui Chen and Zhao (2020), and Dias et al. (2022), shedding light on market fluctuations that classical finance theories could not fully explain. Machine Learning models used in this research demonstrate their ability to predict market volatility, affirming the role of sentiment factors on investor behaviour. These findings contribute to our understanding of the crypto market, offering explanations for its rapid shifts and bubbles, in accordance with the foundational theory of Shiller (2003). This study also confirms that market volatility is affected by investor sentiment, which varies with market conditions, providing empirical evidence supporting the application of these theories in cryptocurrency market analysis.

Our study has multiple practical implications, offering market participants an insight thru a new lens which to view risk management and portfolio allocation in the evolving crypto domain. By understanding the predictive power of sentiment analysis coupled with machine learning, investors and traders can make more informed decisions, potentially leading to improved portfolio performance. For the academic community, our study paves the way for further exploration into the integration of advanced analytics with traditional financial theories. It challenges researchers to think beyond conventional models and explore the rich potential of interdisciplinary approaches in financial analysis.

In conclusion, our study not only contributes to the existing body of knowledge but also opens new avenues for future research. It highlights the growing importance of alternative data sources, such as sentiment data, in financial forecasting and the pivotal role of machine learning in harnessing this data. As the financial world becomes increasingly complex and interconnected, such innovative approaches will be crucial in navigating market dynamics and making sound investment decisions.

References

Abraham, J., Higdon, D., Nelson, J., Ibarra, J., 2018. Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. Technical Report 1 (3). SMU Data Science Review. https://scholar.smu.edu/datasciencereview/vol1/iss3/1.

Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research, 297*, 3–36. https://doi.org/10.1007/s10479-020-03575-y

Alessandretti, L., ElBahrawy, A., Aiello, L. M., & Baronchelli, A. (2018). Anticipating cryptocurrency prices using machine learning. *Complexity, 2018*, 8983590. https://doi.org/10.1155/2018/8983590

Atsalakis, G.S., Atsalaki, I.G., Pasiouras, F., Zopounidis, C., 2019. Bitcoin price forecasting with neuro-fuzzy techniques. European Journal of Operational Research 276, 770–780. https://www.sciencedirect.com/science/article/pii/S037722171930075X, doi:10.1016/j.ejor.2019.01.040.

Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2017). Can volume predict bitcoin returns and volatility? a quantiles-based approach. *Economic Modelling, 64*, 74–81.

Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. Journal of Computational Science 2, 1–8. https://www.sciencedirect.com/science/article/pii/S187775031100007X, doi:10.1016/j.jocs.2010.12.007.

Bouri, E., Gupta, R., 2021. Predicting bitcoin returns: Comparing the roles of newspaper- and internet search-based measures of uncertainty. Finance Research Letters 38, 101398. https://www.sciencedirect.com/science/article/pii/S1544612319307020, doi:10.1016/j.frl.2019.101398.

Bouri, E., Saeed, T., Vo, X.V., Roubaud, D., 2021. Quantile connectedness in the cryptocurrency market. Journal of International Financial Markets, Institutions and Money 71, 101302. https://www.sciencedirect.com/science/article/pii/S1042443121000214, doi:10.1016/j.intfin.2021.101302.

Bouteska, A., Mefteh-Wali, S., Dang, T., 2022. Predictive power of investor sentiment for bitcoin returns: Evidence from covid-19 pandemic. Technological Forecasting and Social Change 184, 121999. https://www.sciencedirect.com/science/article/pii/S0040162522005200, doi:10.1016/j.techfore.2022.121999.

Brauneis, A., Mestel, R., & Theissen, E. (2024). The crypto world trades at tea time: intraday evidence from centralized exchanges across the globe. *Review of Quantitative Finance and Accounting*. https://doi.org/10.1007/s11156-024-01304-1

Bukovina, J., Marticek, M., 2016. Sentiment and Bitcoin Volatility. MENDELU Working Papers in Business and Economics 2016-58. Mendel University in Brno, Faculty of Business and Economics. https://EconPapers.repec.org/RePEc:men:wpaper:58_2016.

Catania, L., Grassi, S., 2022a. Forecasting cryptocurrency volatility. International Journal of Forecasting 38, 878–894. https://www.sciencedirect.com/science/article/pii/S0169207021001059, doi:10.1016/j.ijforecast.2021.06.005.

Catania, L., Grassi, S., 2022b. Forecasting cryptocurrency volatility. International Journal of Forecasting 38, 878–894. https://www.sciencedirect.com/science/article/pii/S0169207021001059, doi:10.1016/j.ijforecast.2021.06.005.

Catania, L., Grassi, S., Ravazzolo, F., 2019. Forecasting cryptocurrencies under model and parameter instability. International Journal of Forecasting 35, 485–501. https://www.sciencedirect.com/science/article/pii/S0169207018301584, doi:10.1016/j.ijforecast.2018.09.005.

Chen, T., Guestrin, C., 2016. XGBoost, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 785–794. , doi:10.1145/2939672.2939785.

Chen, W., Xu, H., Jia, L., Gao, Y., 2021. Machine learning model for bitcoin exchange rate prediction using economic and technology determinants. International Journal of Forecasting 37, 28–43. https://www.sciencedirect.com/science/article/pii/S0169207020300431, doi:10.1016/j.ijforecast.2020.02.008.

Cho, H., Kim, Y., Lee, E., Choi, D., Lee, Y., & Rhee, W. (2020). Basic enhancement strategies when using bayesian optimization for hyperparameter tuning of deep neural networks. *IEEE Access, 8*, 52588–52608. https://doi.org/10.1109/ACCESS.2020.2981072

Conghui Chen, L.L., Zhao, N., 2020. Fear sentiment, uncertainty, and bitcoin price dynamics: The case of covid-19. Emerging Markets Finance and Trade 56, 2298–2309. , doi:10.1080/1540496X.2020.1787150, arXiv:https://doi.org/10.1080/1540496X.2020.1787150.

Corbet, S., Larkin, C., Lucey, B.M., Meegan, A., Yarovaya, L., 2020. The impact of macroeconomic news on bitcoin returns. The European Journal of Finance 26, 1396–1416. , doi:10.1080/1351847X.2020.1737168, arXiv:https://doi.org/10.1080/1351847X.2020.1737168.

Corsi, F., 2009. A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics 7, 174–196. , doi:10.1093/jjfinec/nbp001, arXiv:https://academic.oup.com/jfec/article-pdf/7/2/174/2543795/nbp001.pdf.

Da, Z., Engelberg, J., Gao, P., 2014. The Sum of All FEARS Investor Sentiment and Asset Prices. The Review of Financial Studies 28, 1–32. , doi:10.1093/rfs/hhu072, arXiv:https://academic.oup.com/rfs/article-pdf/28/1/1/2445.0249/hhu072.pdf.

Dias, I.K., Fernando, J.R., Fernando, P.N.D., 2022. Does investor sentiment predict bitcoin return and volatility? a quantile regression approach. International Review of Financial Analysis 84, 102383. https://www.sciencedirect.com/science/article/pii/S1057521922003337, doi:10.1016/j.irfa.2022.102383.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests. *Journal of Business & Economic Statistics, 33*, 1–1. https://doi.org/10.1080/07350015.2014.983236

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 20*, 134–144. https://doi.org/10.1198/073500102753410444

Figá-Talamanca, G., & Patacca, M. (2019). Does market attention affect bitcoin returns and volatility? *Decisions in Economics and Finance, 42*, 135–155. https://doi.org/10.1007/s10203-019-00258-7

Fry, J., & Cheah, E. T. (2016). Negative bubbles and shocks in cryptocurrency markets. *International Review of Financial Analysis, 47*, 343–352.

Ftiti, Z., Louhichi, W., & Ben Ameur, H. (2021). Cryptocurrency volatility forecasting: What can we learn from the first wave of the covid-19 outbreak? *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04116-x

Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to Forget: Continual Prediction with LSTM. Neural Computation 12, 2451–2471. , doi:10.1162/089976600300015015, arXiv:https://direct.mit.edu/neco/article-pdf/12/10/2451/814643/089976600300015015.pdf.

Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market. *Proceedings of the International AAAI Conference on Web and Social Media, 4*, 58–65. https://doi.org/10.1609/icwsm.v4i1.14023

Gradojevic, N., Kukolj, D., Adcock, R., Djakovic, V., 2023. Forecasting bitcoin with technical analysis: A not-so-random forest? International Journal of Forecasting 39, 1–17. https://www.sciencedirect.com/science/article/pii/S0169207021001230, doi:10.1016/j.ijforecast.2021.08.001.

Guégan, D., Renault, T., 2021. Does investor sentiment on social media provide robust information for bitcoin returns predictability? Finance Research Letters 38, 101494. https://www.sciencedirect.com/science/article/pii/S1544612319314199, doi:10.1016/j.frl.2020.101494.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497. https://doi.org/10.3982/ECTA5771

Hoang, L.T., Baur, D.G., 2020. Forecasting bitcoin volatility: Evidence from the options market. Journal of Futures Markets 40, 1584–1602. , doi:10.1002/fut.22144, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/fut.22144.

Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 06, 107–116. , doi:10.1142/S0218488598000094, arXiv:https://doi.org/10.1142/S0218488598000094.

Jana, R. K., Ghosh, I., & Das, D. (2021). A differential evolution-based regression framework for forecasting bitcoin price. *Annals of Operations Research, 306*, 295–320. https://doi.org/10.1007/s10479-021-04000-8

Jo, H., Park, H., Shefrin, H., 2020. Bitcoin and sentiment. Journal of Futures Markets 40, 1861–1879. https://onlinelibrary.wiley.com/doi/abs/10.1002/fut.22156, doi:10.1002/fut.22156, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/fut.22156.

Karalevicius, V., Degrande, N., & De Weerdt, J. (2018). Using sentiment analysis to predict interday bitcoin price movements. *The Journal of Risk Finance, 19*, 56–75. https://doi.org/10.1108/JRF-06-2017-0092

Katsiampa, P., 2017. Volatility estimation for bitcoin: A comparison of garch models. Economics Letters 158, 3–6. https://www.sciencedirect.com/science/article/pii/S0165176517302501, doi:10.1016/j.econlet.2017.06.023.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 1–9. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Kraaijeveld, O., De Smedt, J., 2020. The predictive power of public twitter sentiment for forecasting cryptocurrency prices. Journal of International Financial Markets, Institutions and Money 65, 101188. https://www.sciencedirect.com/science/article/pii/S104244312030072X, doi:10.1016/j.intfin.2020.101188.

Krauss, C., Do, X.A., Huck, N., 2017. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s &p 500. European Journal of Operational Research 259, 689–702. https://www.sciencedirect.com/science/article/pii/S0377221716308657, doi:10.1016/j.ejor.2016.10.031.

Kristoufek, L., 2013. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. Scientific Reports 3, 3415. https://doi.org/10.1038/srep03415, doi:10.1038/srep03415.

Kumar Kulbhaskar, A., Subramaniam, S., 2023. Breaking news headlines: Impact on trading activity in the cryptocurrency market. Economic Modelling 126, 106397. https://www.sciencedirect.com/science/article/pii/S0264999323002092, doi:10.1016/j.econmod.2023.106397.

Kyriazis, N., Papadamou, S., Tzeremes, P., Corbet, S., 2022. The differential influence of social media sentiment on cryptocurrency returns and volatility during covid-19. The Quarterly Review of Economics and Finance https://www.sciencedirect.com/science/article/pii/S1062976922001041, doi:10.1016/j.qref.2022.09.004.

Köchling, G., Schmidtke, P., Posch, P.N., 2020. Volatility forecasting accuracy for bitcoin. Economics Letters 191, 108836. https://www.sciencedirect.com/science/article/pii/S0165176519304239, doi:10.1016/j.econlet.2019.108836.

Lin, Y., Yan, Y., Xu, J., Liao, Y., Ma, F., 2021. Forecasting stock index price using the ceemdan-lstm model. The North American Journal of Economics and Finance 57, 101421. https://www.sciencedirect.com/science/article/pii/S1062940821000553, doi:10.1016/j.najef.2021.101421.

Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. Journal of Econometrics 187, 293–311. https://www.sciencedirect.com/science/article/pii/S0304407615000329, doi:10.1016/j.jeconom.2015.02.008.

Liu, M., Li, G., Li, J., Zhu, X., Yao, Y., 2021. Forecasting the price of bitcoin using deep learning. Finance Research Letters 40, 101755. https://www.sciencedirect.com/science/article/pii/S1544612320304864, doi:10.1016/j.frl.2020.101755.

Ángeles López-Cabarcos, M., Pérez-Pico, A.M., Piñeiro-Chousa, J., Šević, A., 2021. Bitcoin volatility, stock market and investor sentiment. are they connected? Finance Research Letters 38, 101399. https://www.sciencedirect.com/science/article/pii/S1544612319309274, doi:10.1016/j.frl.2019.101399.

Mallqui, D.C., Fernandes, R.A., 2019. Predicting the direction, maximum, minimum and closing prices of daily bitcoin exchange rate using machine learning techniques. Applied Soft Computing 75, 596–606. https://www.sciencedirect.com/science/article/pii/S1568494618306707, doi:10.1016/j.asoc.2018.11.038.

McNally, S., Roche, J., Caton, S., 2018. Predicting the price of bitcoin using machine learning, in: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), pp. 339–343. doi:10.1109/PDP2018.2018.00060.

Moser, S., Brauneis, A., 2023. Should you listen to crypto youtubers? Finance Research Letters 54, 103782. https://www.sciencedirect.com/science/article/pii/S1544612323001551, doi:10.1016/j.frl.2023.103782.

Naeem, M.A., Mbarki, I., Shahzad, S.J.H., 2021. Predictive role of online investor sentiment for cryptocurrency market: Evidence from happiness and fears. International Review of Economics & Finance 73, 496–514. https://www.sciencedirect.com/science/article/pii/S1059056021000083, doi:10.1016/j.iref.2021.01.008.

Panagiotidis, T., Stengos, T., Vravosinos, O., 2019. The effects of markets, uncertainty and search intensity on bitcoin returns. International Review of Financial Analysis 63, 220–242. https://www.sciencedirect.com/science/article/pii/S105752191830499X, doi:10.1016/j.irfa.2018.11.002.

Peng, S., Prentice, C., Shams, S., Sarker, T., 2023. A systematic literature review on the determinants of cryptocurrency pricing. China Accounting and Finance Review ahead-of-print. https://doi.org/10.1108/CAFR-05-2023-0053, doi:10.1108/CAFR-05-2023-0053.

Plakandaras, V., Bouri, E., & Gupta, R. (2021). Forecasting bitcoin returns: is there a role for the us-china trade war? *Journal of Risk, 23*, 75–93. https://doi.org/10.21314/JOR.2021.001

Qiu, Y., Wang, Y., Xie, T., 2021. Forecasting bitcoin realized volatility by measuring the spillover effect among cryptocurrencies. Economics Letters 208, 110092. https://www.sciencedirect.com/science/article/pii/S0165176521003694, doi:10.1016/j.econlet.2021.110092.

Rognone, L., Hyde, S., Zhang, S.S., 2020. News sentiment in the cryptocurrency market: An empirical comparison with forex. International Review of Financial Analysis 69, 101462. https://www.sciencedirect.com/science/article/pii/S105752192030106X, doi:10.1016/j.irfa.2020.101462.

Sahiner, M., McMillan, D. G., & Kambouroudis, D. (2023). Do artificial neural networks provide improved volatility forecasts: Evidence from asian markets. *Journal of Economics and Finance, 47*, 723–762. https://doi.org/10.1007/s12197-023-09629-8

Sapkota, N., 2022. News-based sentiment and bitcoin volatility. International Review of Financial Analysis 82, 102183. https://www.sciencedirect.com/science/article/pii/S1057521922001454, doi:10.1016/j.irfa.2022.102183.

Sermpinis, G., Karathanasopoulos, A., Rosillo, R., & de la Fuente, D. (2021). Neural networks in financial trading. *Annals of Operations Research, 297*, 293–308. https://doi.org/10.1007/s10479-019-03144-y

Shen, D., Urquhart, A., Wang, P., 2019. Does twitter predict bitcoin? Economics Letters 174, 118–122. https://www.sciencedirect.com/science/article/pii/S0165176518304634, doi:10.1016/j.econlet.2018.11.007.

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives, 17*, 83–104. https://doi.org/10.1257/089533003321164967

Sun, X., Liu, M., Sima, Z., 2020. A novel cryptocurrency price trend forecasting model based on lightgbm. Finance Research Letters 32, 101084. https://www.sciencedirect.com/science/article/pii/S1544612318307918, doi:10.1016/j.frl.2018.12.032.

Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance 62, 1139–1168. doi:10.1111/j.1540-6261.2007.01232.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2007.01232.x.

Trucís, C., 2019. Forecasting bitcoin risk measures: A robust approach. International Journal of Forecasting 35, 836–847. https://www.sciencedirect.com/science/article/pii/S0169207019300184, doi:10.1016/j.ijforecast.2019.01.003. forecasting issues in developing economies.

Walther, T., Klein, T., Bouri, E., 2019. Exogenous drivers of bitcoin and cryptocurrency volatility - a mixed data sampling approach to forecasting. Journal of International Financial Markets, Institutions and Money 63, 101133. https://www.sciencedirect.com/science/article/pii/S1042443119302446, doi:10.1016/j.intfin.2019.101133.

Wang, J., Ma, F., Bouri, E., Guo, Y., 2023. Which factors drive bitcoin volatility: Macroeconomic, technical, or both? Journal of Forecasting https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2930, doi:10.1002/for.2930, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.2930.

Wei, M., Sermpinis, G., Stasinakis, C., 2023. Forecasting and trading bitcoin with machine learning techniques and a hybrid volatility/sentiment leverage. Journal of Forecasting 42, 852–871. https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2922, doi:10.1002/for.2922, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.2922.

Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., Deng, S.H., 2019. Hyperparameter optimization for machine learning models based on bayesian optimizationb. Journal of Electronic Science and Technology 17, 26–40. https://www.sciencedirect.com/science/article/pii/S1674862X19300047, doi:10.11989/JEST.1674-862X.80904120.

Xia, Y., Sang, C., He, L., Wang, Z., 2023. The role of uncertainty index in forecasting volatility of bitcoin: Fresh evidence from garch-midas approach. Finance Research Letters 52, 103391. https://www.sciencedirect.com/science/article/pii/S1544612322005682, doi:10.1016/j.frl.2022.103391.