Benchmarking Modeling Architectures for Cryptocurrency Price prediction using Financial and Social Media Data

Tulika Shrivastava
The University of Sydney
Basem Suleiman
The University of Sydney
Sachit Desa
sachit.desa@machester.ac.uk
The University of Sydney
Muhammad Johan Alibasa
Monash University
Andrey Tregubov
The University of Sydney
Jiewen Zhang
The University of Sydney

Research Article

Keywords: Cryptocurrency price prediction, Hybrid models, Sentiment analysis, Financial data, Social media data

Posted Date: April 21st, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6288160/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Social Network Analysis and Mining on August 30th, 2025. See the published version at https://doi.org/10.1007/s13278-025-01520-0.

Benchmarking Modeling Architectures for Cryptocurrency Price prediction using Financial and Social Media Data

Tulika Shrivastava[1], Basem Suleiman[1,2], Sachit Desa[1,3*], Muhammad Johan Alibasa[4], Andrey Tregubov[1], Jiewen Zhang[1]

[1]School of Computer Science, The University of Sydney, Camperdown, Sydney, 2050, NSW, Australia.
[2]School of Computer Science, The University of New South Wales, , Sydney, 2052, NSW, Australia.
[3*]School of Environment, Education and Development, The University of Manchester, M13 9PL,United Kingdom.
[4]School of Computer Science, Monash University, , Melbourne, 3800, Vic, Australia.

*Corresponding author(s). E-mail(s): sachit.desa@machester.ac.uk;
Contributing authors: tshr3215@uni.sydney.edu.au; b.suleiman@unsw.edu.au; johan.alibasa@monash.edu; andrey.tregubov@sydney.edu.au; jiewen.zhang@sydney.edu.au;

Abstract

The volatility of cryptocurrencies necessitates reliable short-term price prediction models for informed investment decisions. This work presents two benchmarking studies that predict cryptocurrency price over hourly and daily time horizons based on market indicators and social media data. The studies evaluate various models, network architectures, and feature combinations, with a focus on understanding the influence of lag periods, data aggregation, and sentiment analysis nuances on cryptocurrency price. Empirical evaluation identifies LSTM to be the leading prediction model, outperforming other singular models including CNN, ARIMA, and SVR. Hybrid architectures combining LSTM with statistical methods, ARIMA or ARIMAX, demonstrate robust performance over a range of model configurations and combinations. Overall, the findings of this study offer practical contributions on how statistical methods combined with deep learning can address non-stationarity, preprocess social media data, and help support feature interpretability for cryptocurrency price prediction.

Keywords: Cryptocurrency price prediction, Hybrid models, Sentiment analysis, Financial data, Social media data

1 Introduction

Recent years have seen a boom in cryptocurrency investment. Generally, a cryptocurrency maybe defined as a digital peer-to-peer electronic payment system, in which transaction records are maintained by a decentralized system via cryptography. The popularity of this currency form is driven by: (i) the security of cryptographic transactions, which make fraud and hacking difficult; (ii) absence of a controlling entity, such as a bank or financial institution; and (iii) global accessibility, which ensures that these assets are traded globally without incurring fees or levies. However, despite their appeal, cryptocurrencies are still viewed as a speculative investment. This is in-part a result of their round-the-clock trading which adds to their price volatility. Due to this unpredictable pricing nature, a strong forecasting model is a welcome tool for investors considering cryptocurrency trading, especially over the short-term.

Kristoufek [1] argues that the fundamental price of cryptocurrency is driven by their *'usage in trade, money supply, and price level'*. Other studies have extended this perspective by leveraging market metrics for forecasting price fluctuations [2]. A separate stream of work, often interlinked with price forecasting, has investigated the role of social media sentiment as a price driver of cryptocurrencies. However, existing studies often limit their scope to selective algorithms, feature sets, or simplified sentiment processing approaches, leaving critical gaps in understanding of the interplay between market dynamics and sentiment-driven behavior.

This work bridges these gaps by presenting the results of two short-term cryptocurrency price prediction studies that base prediction on market indicators and social media data. Both studies adopt common predictive baseline models such as ARIMA, CNN, and LSTM. The studies explore advanced hybrid architectures and diverse feature selection techniques, offering valuable perspectives on the application of machine learning and deep learning for financial time-series forecasting. The findings are also relevant to broader time-series forecasting scenarios involving multi-format data. The specific research contributions of this work are as follows:

- Identifying singular and hybrid model architectures that optimize forecasting of financial time-series data.
- Evaluating the effects of social media data preprocessing and feature selection on cryptocurrency price prediction accuracy.
- Benchmarking the performance of combined model architectures on multi-format datasets.

2 Related Work

Numerous studies have sought to predict cryptocurrency prices using market indicators and social media data. This section reviews the state of research on singular prediction models, network architectures, and the role of social media sentiment in cryptocurrency price forecasting. Also highlighted are the key knowledge gaps across these areas and in our understanding of the factors driving short-term cryptocurrency price movements.

2.1 Singular Price Prediction Models

Statistical models like ARIMA have been widely used for short-term cryptocurrency price forecasting. Hua [3] highlights ARIMA's effectiveness in short-term predictions but notes its limitations for long-term horizons. Conversely, Support Vector Regression (SVR), a non-parametric model, is adept at capturing non-linear, non-stationary processes, producing competitive results against deep learning methods [2].

Deep learning models such as 1D-CNN and LSTM have demonstrated superior performance for price prediction [2, 4–7]. LSTM is particularly effective at modeling long-term dependencies due to its gated architecture, enabling it to capture significant historical events and their influence on current prices. CNNs have also been utilized for automatic feature extraction and modeling non-linear relationships [6, 8]. Together, these models set a baseline for forecasting accuracy, which more advanced architectures aim to surpass.

2.2 Network Architectures

Several network architectures show promise in the short-term price prediction. Ensemble models leveraging LSTM or CNN as base learners often outperform singular models by mitigating biases in individual classifiers. Ji et al. [6] demonstrate the consistent superiority of ensemble architectures over traditional methods. Multi-modal models, which assign specific feature subsets to tailored base learners, further improve forecasting accuracy [2, 9, 10]. Hybrid architectures that combine statistical methods (like ARIMA) with neural networks (such as RNN) leverage the complementary strengths of both approaches. Statistical models effectively capture linear patterns and seasonality, while neural networks handle complex non-linear relationships. When properly configured, hybrid models demonstrate superior predictive performance and robustness [11].

Transformers have recently gained attention for time-series forecasting. Coupled with LSTM, they have shown potential in improving accuracy [12–14]. However, their computational demands and inconsistent performance for short-term horizons make them less practical for real-time cryptocurrency predictions. Specifically, their computational complexity acts as a bottleneck for real-time applications requiring fast predictions [15]. While combined architectures generally exhibit superior predictive accuracy, challenges remain in configuring these models, processing data effectively, and applying them to the specific context of multi-modal data for short-term price prediction [14].

2.3 Social Media Sentiment

Market indicators such as price highs, lows, and trade volume are established drivers of cryptocurrency price [1, 2]. The influence of social media sentiment, however, is less definitive. While Bitcoin price has been shown to correlate with sentiment extremes and tweet volume, the overall impact of sentiment is debated. Abraham et al. [10] argue that tweet volume is a stronger predictor than sentiment, recommending more sophisticated models to explore this relationship in volatile markets. Wołk [16] and Youssfi et al. [17] expand this analysis to include Google Trends, demonstrating the predictive power of search interest and negative sentiment in short-term forecasting.

LSTM consistently demonstrates superior performance in cryptocurrency price prediction compared to other architectures. However, optimal configurations for deploying LSTM and CNN remain unclear, particularly regarding model parameterization and prediction horizons. The integration of social media data into these architectures presents additional challenges, specifically in distinguishing significant market events from general sentiment fluctuations. A more robust assessment of temporal factors - particularly how prediction horizons influence data aggregation methods - is needed to better understand their effects on model performance and feature importance.

This work advances the understanding of multi-modal architectures for short-term cryptocurrency price prediction through empirical evaluation of both market indicators and social media features. The comprehensive benchmarking of model performance provides evidence-based insights into the effective integration of these data.

3 Data Collection and Processing

This section describes the data used in both studies, the preprocessing methods applied to social media data, and the transformation of financial indicators. Study 1 utilized BERT for sentiment analysis, while Study 2 employed VADER. Both sentiment analysis models were fine-tuned on cryptocurrency-specific tweets to capture relevant nuances and colloquialisms before deployment. Study 1 examined price prediction over hourly intervals, while Study 2 focused on daily forecasting horizons. Financial and social media data were aggregated to match these respective prediction windows.

3.1 Study 1: Sentiment Inference model and market indicators

Twitter data used for study 1 was sourced from Kaggle and covered a time period of 3 years, from January 2017 to December 2019 [18]. The data was made available in its raw format with information relating usernames, urls, timestamps, likes, retweets, etc...The dataset consisted of a total of 16M tweets.

In the interest of selecting an appropriate sentiment analysis model, three inference models were evaluated for their accuracy: VADER (rule-based), TextBlob (machine learning), and BERT (deep learning). The models were evaluated over labeled twitter data, with BERT performing notably better across accuracy, precision, recall, and F1-scores. BERT was selected as the preferred sentiment analysis model and generated a sentiment label (positive, neutral, or negative) for each tweet in the dataset. The labels were categorically aggregated on an hourly basis and matched with the time increment and cryptocurrency price (hourly representation of market data). The number of likes for each tweet were counted towards the classified sentiment label of that tweet. For instance, if a tweet had 100 likes, and it was classified positive, then a count of 101 was added to the total number of positive opinions. The final twitter dataset consisted of: timestamp (hourly), number of positive opinions, negative opinions, and neutral opinions.

Cryptocurrency market data, recorded at 1-minute intervals, was sourced from Kaggle for the same period as the Twitter data. This dataset included details such as date, opening price, closing price, high price, low price, and transaction volume. Data was aggregated to hourly intervals, retaining the timestamp, highest price, lowest price, transaction volume, and closing price for the next time step. Figure 1 presents a candlestick chart showing Bitcoin's opening price and its correlation with Twitter sentiment counts.

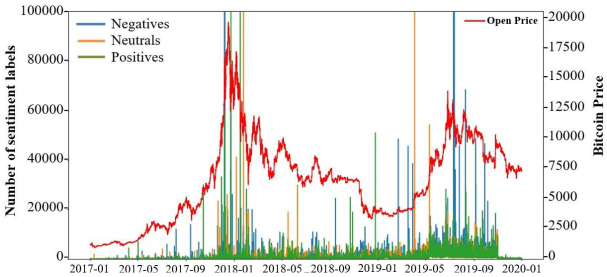

Fig. 1: Candlestick chart for Bitcoin price from January 2017 to December 2019.

3.2 Study 2: Sentiment weighting, Google Trends, and market indicators

In Study 2, Twitter data related to Bitcoin and Ethereum was collected over three years via API. Data preprocessing involved removing extraneous elements such as hashtags, URLs, and mentions while retaining text-based information. Automated accounts (algorithmically-driven profiles or "bots") were also excluded from the dataset, consistent with study 1. VADER was chosen for sentiment analysis in this study due to its lightweight architecture and ability to generate both sentiment labels and compound scores, representing overall sentiment. VADER assigns a positive, neutral, or negative label to each tweet, as well as calculating a compound score. The score represents the overall sentiment of a string of text (i.e. sum of individual word scores). To derive the daily sentiment score, a weighted average method was employed. The weighted average took into account the metadata for each tweet: number of replies, likes, quotes, followers, accounts followed, and total number of tweets. To avoid the skew introduced by influential accounts, the metadata was normalised before the weighted average was calculated. Eq (1) presents the calculation of daily sentiment score.

$$\text{sentiment score} \rightarrow \frac{\sum(\text{compound score}_{\text{norm}} \times \text{weight}_{\text{norm}})}{\sum(\text{weight}_{\text{norm}})} \tag{1}$$

Google Trends data (January 2019–March 2022) was included to capture a broader spectrum of investor sentiment, accounting for individuals not actively engaged on Twitter. Features such as search interest and related topics were transformed into count data and added to the social media dataset. Specifically, Google Trends data was transformed into count data using a threshold-based conversion method. Search interest values (0-100) were mapped to count values based on intensity levels, with higher search interest corresponding to higher count values. This transformation preserved the relative importance of search trends while providing a more concrete representation of public interest. The resulting count data was then normalized to account for overall search volume variations across different time periods.

Financial market data related to Bitcoin and Ethereum were sourced from [19]. The data consisted of opening price, closing price, high, and low market values, and did not require any preprocessing.

3.3 Evaluating Lag-times

A comprehensive analysis of lag structures for Bitcoin and Ethereum was conducted using multiple statistical approaches. The temporal relationships between cryptocurrency prices and social media sentiment were evaluated through three complementary methods: Granger causality tests to establish directional relationships, cross-correlation analysis to identify temporal dependencies, and Autoregressive Distributed Lag (ARDL) modeling to capture dynamic interactions. This multi-method approach provided robust validation of the optimal lag structures across both studies. The empirical analysis revealed distinct temporal patterns for each cryptocurrency. For Bitcoin, Granger causality tests (p ¡ 0.05) demonstrated that tweet count significantly influences closing price after a single lag period. In contrast, Ethereum exhibited a longer response time, with statistical significance emerging at 5 lags. These findings were corroborated by cross-correlation analysis, which showed peak correlations at corresponding lag periods. The temporal relationship between Bitcoin's closing price and tweet volume is visualized in Figure 2 which presents the results of Granger causality.

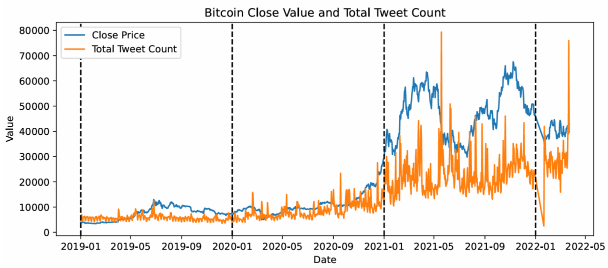

Fig. 2: Granger causality for Bitcoin close price and tweet volume

Table 1 summarizes the experimental configurations for both studies, summarizing the sentiment analysis models, data sources, and temporal parameters.

4 Methodology

In both studies, 1D-CNN and LSTM models were employed as individual predictors to establish baseline accuracies for cryptocurrency price prediction. Building upon these baselines, more sophisticated modeling architectures were developed to enhance predictive performance and surpass the benchmarks set by individual models. The experimental framework, outlined in Tables 2 & 3, systematically examines price prediction

Table 1: Experimental Configuration Across Studies

Feature	**Study 1**	**Study 2**
Sentiment Analysis Model	BERT	VADER
Social Media Data	Twitter	Twitter & Google Trends
Market Data	Bitcoin	Bitcoin & Ethereum
Lag Period	1 lag	1 lag (Bitcoin) 5 lags (Ethereum)
Prediction Horizon	1-hour	1-day

across model architectures, feature combinations, and feature processing techniques.

The decision to predict cryptocurrency prices rather than *returns* was driven by the specific goals of this research. While financial modeling often favors *returns* due to their stationarity, price prediction aligns more closely with the practical needs of traders and market participants, who often prioritize forecasting absolute price levels for decision-making. Returns, defined as percentage changes in price over time, remove trends and seasonality, making them ideal for econometric models. However, the current approach in both studies addresses non-stationarity in price prediction through established preprocessing techniques. In doing so, both studies maintain interpretability of direct price predictions. Furthermore, by leveraging deep learning models (1D-CNN and LSTM), which are more flexible in handling varied data characteristics, both studies were able to model complex temporal dependencies directly from price data.

4.1 Baseline Models

The **A**uto**R**egressive, **I**ntegrated, **M**oving **A**verage (ARIMA) is a widely used statistical time series prediction model. ARIMA operates under the assumption of stationarity, where statistical properties like mean and variance remain constant over time. To satisfy this assumption, the input time series were tested for stationarity using the Augmented Dickey-Fuller (ADF) test and transformed via log and first-order differencing. For cases involving additional independent variables, ARIMAX was used, allowing the inclusion of external factors such as sentiment or trend data. This extension is particularly beneficial for capturing relationships between market dynamics and external influences. The parameters, q, representing current values based on past residual errors, and, p, representing lagged observations, were tuned via a parameter grid search using the Akaike Information Criterion (AIC).

The choice of Support Vector Machine (SVR) in study 1 was inspired by it's effectiveness in capturing non-linear relationships to successfully predict Bitcoin price [20]. The parameters tuned for SVR were: a regularization parameter (C), margin separation parameter (ε), and width of the radial basis function (RBF) kernel (γ). Optimal values for these parameters were identified through k-fold cross-validation to prevent over-fitting and ensure robust performance.

The LSTM models used in this study required the following hyperparameters to be tuned: prediction window size, number of hidden units in the LSTM block, and batch size. For the 1D-CNN model, the parameters to be tuned included window size, number of conv-1d blocks, and batch size. Hyperparameter tuning was performed through randomized search on a predefined parameter grid, with evaluation based on R^2, RMSE, and MAPE metrics. For robust training, and in order to prevent information leakage, the following measures were implemented: a. Testing and training data were split chronologically to ensure future data did not influence past predictions; b. Transformations were performed on the training data and these same transformations replicated on the validation and test sets using parameters derived from training sets; and c. Temporal cross-validation was used to ensure that validation folds respected time-series structure.

4.2 Study 1: Ensemble and Multi-modal models

Bespoke architectures built in study 1 include an ensemble model and multi-modal model. Both modeling architectures employed 1D-CNNs and LSTMs as base learners. Since the hyperparameters of 1D CNN and the LSTM model were tuned separately, only parameters related to optimal window size and batch size were tuned for the ensemble and multi-modal models. Tuning the window and batch size allowed for the optimization of model performance by controlling the trade-off between computational time and accuracy.

The ensemble model consists of two base-learners, CNN and LSTM, which accept the same input data. Each base learner generates its own output, that is concatenated and fed into a dense layer. This layer optimizes the final prediction by calculating a weighted combination of individual predictions. Figure 3 illustrates the architecture. The multi-modal model, differing only in input modalities, separates data streams such that one base learner receives market indicators while the other processes sentiment data. The experimental setup for study 1 is presented in Table 2.

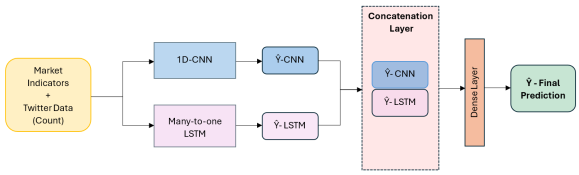

Fig. 3: Ensemble model architecture

Table 2: Experimental setup for study 1

Method	**Algorithm**	**Dataset Features**
Singular	ARIMA	Market indicators
	SVR	Market indicators
		Market indicators+Twitter
	CNN	Market indicators
		Market indicators+Twitter
	LSTM	Market indicators
		Market indicators+Twitter
Ensemble	CNN+LSTM	Market indicators
		Market indicators + Twitter
Multimodal	CNN+LSTM	Market(LSTM)+Twitter (CNN)
		Market(CNN)+Twitter (LSTM)

4.3 Study 2: Hybrid models

Study 2 employs a hybrid modeling approach that combines deep learning models (LSTM or 1D-CNN) with statistical methods (ARIMA or ARIMAX). For financial data analysis, LSTM and 1D-CNN were integrated with ARIMA, while analysis incorporating both financial and non-financial data utilized ARIMAX. The hybrid architecture processes data in two stages: first, the statistical model generates residuals, which then serve as inputs for the deep learning model. The final predictions are transformed back to the original scale for comparison with test data. This two-stage approach leverages the complementary strengths of both modeling paradigms: statistical methods handle the underlying data structure while deep learning captures complex patterns in the residuals. Through this combination, the hybrid architecture aims to achieve more robust predictions than either approach alone. Figure 4 illustrates this workflow.

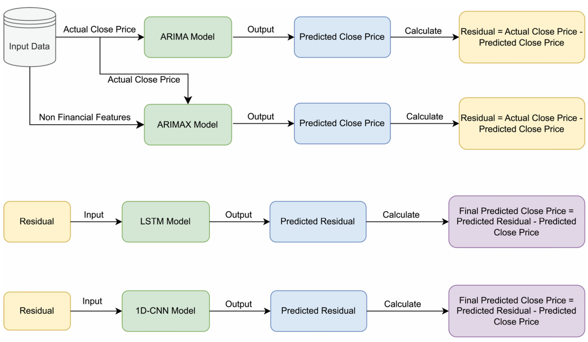

Fig. 4: Hybrid model workflow

Study 2 evaluated four primary feature categories: market indicators (MI), Twitter sentiment (TS), Tweet count (TC), and Google trends (GT). These features were tested in multiple combinations to assess their relative contributions to prediction accuracy and to understand feature interactions. Table 3 details the experimental configurations used to systematically evaluate these feature combinations.

Table 3: Experimental setup for study 2

Method	Algorithm	Dataset Features
Singular	CNN/LSTM	Market Indicators
		Market Indicators, TC
		Market Indicators, TS
		Market Indicators, GT
		Market Indicators, TS, TC & GT
Hybrid	CNN + ARIMA	Market Indicators
	LSTM + ARIMA	Market Indicators
	CNN + ARIMAX	Market Indicators, TC
		Market Indicators, TS
		Market Indicators, GT
		Market Indicators, TS, TC & GT
	LSTM + ARIMAX	Market Indicators, TC
		Market Indicators, TS
		Market Indicators, GT
		Market Indicators, TS, TC & GT

5 Results

Prediction results were evaluated over metrics of MAPE and RMSE, these values have been standardized via Z-scores for comparison. The models for study 1 were trained on Bitcoin data from January, 2017 to August, 2019 and Bitcoin price prediction was completed for the period of September to December of 2019 (4 months). The models for study 2 were trained on cryptocurrency data from January, 2019 to February, 2021 and prediction was complete between February, 2021 to August, 2022. A comparison of actual and predicted prices for both studies are presented in Tables 4 to 6.

Some salient observations based on the results of both studies are listed below :

- Multi-modal, ensemble, and hybrid architecture models demonstrate robust performance, generally outperforming singular models.
- LSTM is the leading time-series prediction model amongst singular models.
- Combined model architectures, with LSTM as a base-learner, demonstrate superior results to networks without LSTM as a base-learner.
- Market indicators, alone, are strong price predictors, even without social media data.

Table 4: Bitcoin Price Prediction for Study 1

Method	Algorithm	Data Features	RMSE	MAPE	Zscore	
					RMSE	MAPE
Singular Models	LSTM	MI	4.56e-03	2.10e-05	-1.09	-1.39
Multi-modal Models	CNN + LSTM	MI (LSTM) + TC (CNN)	4.63e-03	2.10e-05	-1.08	-1.39
Singular Models	LSTM	MI + TC	4.68e-03	2.20e-05	-1.07	-1.38
Multi-modal Models	CNN + LSTM	MI (CNN) + TC (LSTM)	1.28e-02	1.69e-04	-0.06	0.30
Singular Models	CNN	MI + TC	1.30e-02	1.69e-04	-0.05	0.30
Ensemble Models	CNN + LSTM	MI + TC	1.30e-02	1.69e-04	-0.05	0.30
Ensemble Models	CNN + LSTM	MI	1.36e-02	1.86e-04	0.03	0.49
Singular Models	CNN	MI	1.39e-02	1.93e-04	0.07	0.57
Singular Models	SVR	MI	2.53e-02	2.31e-04	1.47	1.01
Singular Models	SVR	MI + TC	2.81e-02	2.46e-04	1.82	1.18
Singular Models	ARIMA	MI	4.00e+03	4.68e+01	–	–

Table 5: Bitcoin Price Prediction for Study 2

Method	Algorithm	Data Features	RMSE	MAPE	Zscore RMSE	Zscore MAPE
Hybrid	LSTM + ARIMA	MI	1725.3	0.027	-1.08	-1.11
	LSTM + ARIMAX	MI, TS, TC & GT	1866.9	0.028	-0.96	-1.07
	LSTM + ARIMAX	MI, TC	1901.5	0.028	-0.93	-1.07
	LSTM + ARIMAX	MI, GT	1903.3	0.028	-0.93	-1.07
	LSTM + ARIMAX	MI, TS	1921.1	0.058	-0.92	0.26
	CNN + ARIMAX	MI, TS, TC & GT	2026.2	0.033	-0.83	-0.85
	CNN + ARIMAX	MI, TS	2066.4	0.033	-0.80	-0.85
	CNN + ARIMAX	MI, TC	2090.6	0.033	-0.78	-0.85
	CNN + ARIMAX	MI, GT	2112.4	0.034	-0.76	-0.80
Singular Models	LSTM	MI	2557.9	0.044	-0.38	-0.36
Singular Models	LSTM	MI, GT	2926.6	0.052	-0.07	-0.01
Hybrid	CNN + ARIMA	MI	2983.2	0.051	-0.02	-0.05
Singular Models	LSTM	MI, TS	3387.5	0.058	0.32	0.26
Singular Models	CNN	MI	3679.9	0.057	0.57	0.21
Singular Models	LSTM	MI, TS, TC & GT	3801.8	0.066	0.67	0.61
Singular Models	LSTM	MI, TC	4075.7	0.075	0.90	1.00
Singular Models	CNN	MI, TC	4151.6	0.070	0.96	0.78
Singular Models	CNN	MI, TS	4520.3	0.077	1.27	1.09
Singular Models	CNN	MI, TS, TC & GT	4742.0	0.083	1.46	1.36
Singular Models	CNN	MI, GT	5743.1	0.109	2.31	2.50

Figures 5 to 7 present a visual comparison of the actual price versus the predicted price for the two best performing models in studies 1 & 2.

The range of Bitcoin price during testing in study 1 ranged from 6,600 to 10,400 USD across 4 months. In contrast, for study 2, the Bitcoin price fluctuated between 39,600 and 67,000 USD, whilst the Ethereum price ranged between 2,300 to 4,900 USD across 16 months. From figure 5 to 7 we infer:

- While Study 2 better predicts upturns and downturns, as shown by the predicted line (red) aligning with the actual price (green) at peaks and troughs, latent patterns and/or other factors still influence these inflections.
- Price prediction in study 2 is accurate for steep increases, decreases, and sudden price inflections, reinforcing the robustness of hybrid architectures.
- Both studies develop models that are robust across a broad pricing range.
- Micro and macros price fluctuations are generally well-captured across the models in both studies.

Table 6: Ethereum Price Prediction for Study 2

Method	**Algorithm**	**Data Features**	**RMSE**	**MAPE**	**Zscore RMSE**	**Zscore MAPE**
Hybrid	LSTM + ARIMAX	MI, TS, TC & GT	153.2	0.033	-0.61	-0.61
	LSTM + ARIMAX	MI, GT	153.6	0.033	-0.60	-0.61
	LSTM + ARIMAX	MI, TC	153.7	0.033	-0.60	-0.61
	LSTM + ARIMAX	MI, TS	154.4	0.033	-0.60	-0.61
Hybrid	CNN + ARIMAX	MI, TS, TC & GT	156.5	0.034	-0.60	-0.60
	CNN + ARIMAX	MI, GT	157.1	0.034	-0.60	-0.60
	CNN + ARIMAX	MI, TS	158.7	0.034	-0.60	-0.60
	CNN + ARIMAX	MI, TC	161.4	0.035	-0.59	-0.59
Hybrid	LSTM + ARIMA	MI	177.6	0.043	-0.56	-0.53
Singular Models	LSTM	MI	212.4	0.051	-0.49	-0.47
	CNN	MI, TS	236.3	0.055	-0.44	-0.43
	CNN	MI	262.3	0.065	-0.39	-0.35
	CNN	MI, TC	284.9	0.066	-0.34	-0.35
	CNN	MI, TS, TC & GT	314.6	0.073	-0.28	-0.29
Singular Models	LSTM	MI, TS, TC & GT	1293.1	0.306	1.66	1.59
	LSTM	MI, TS	1402.9	0.342	1.88	1.88
	LSTM	MI, TC	1495.9	0.370	2.07	2.10
	LSTM	MI, GT	1504.1	0.373	2.08	2.12

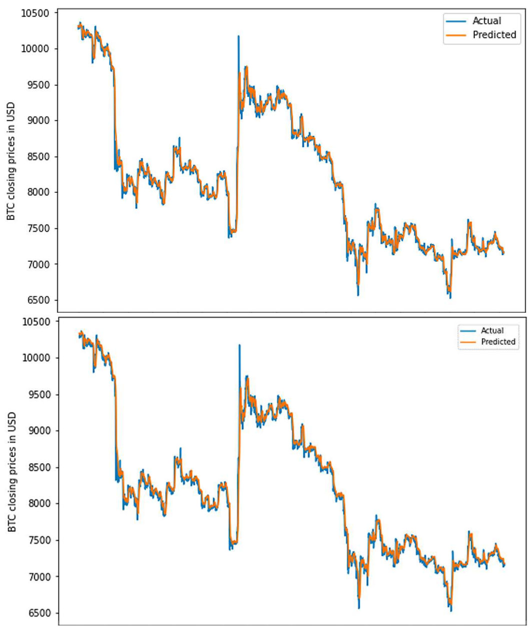

Fig. 5: Two best performing models of study 1: (Up) LSTM with market indicators, (Down) multi-modal model of LSTM (market indicators) and CNN (twitter count)

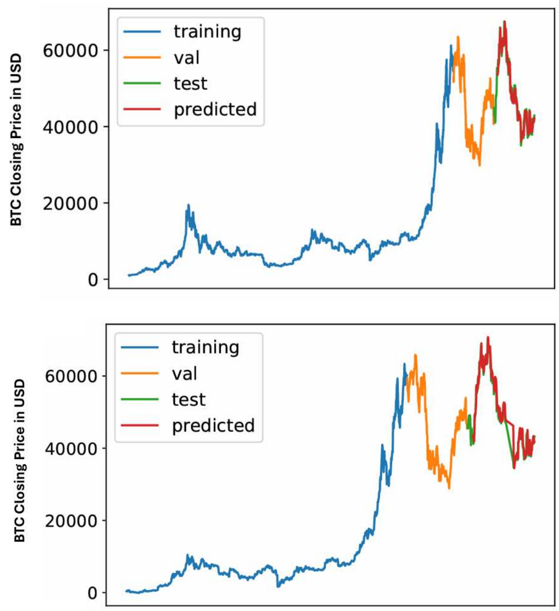

Fig. 6: Two best performing models for Bitcoin in study 2: (Up) Hybrid model of LSTM+ARIMA (market indicators), (Down) Hybrid model of LSTM+ARIMAX (market indicators, twitter count, sentiment, and google trends)

6 Discussion

6.1 Efficiency of Network Architectures

Network architectures consisting of two base learners generally outperform singular models,

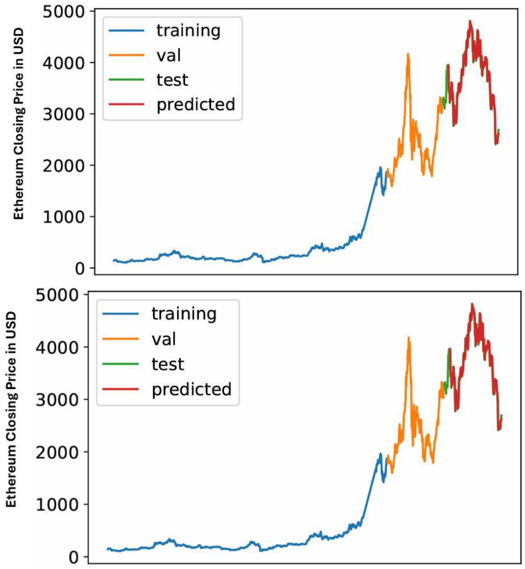

Fig. 7: Two best performing models for Ethereum in study 2: (Up) Hybrid model of LSTM+ARIMAX (market indicators, twitter count, sentiment, and google trends), (Down) Hybrid model of LSTM+ARIMAX (market indicators and google trends)

with the exception of LSTM in study 1. While multi-modal models achieve better accuracy than ensemble models, hybrid models consistently rank as the best predictors.

In study 1, the multi-modal and ensemble architectures combine two deep-learning models, LSTM and CNN. In contrast, the hybrid model architecture leverages the strengths of statistical models, ARIMA or ARIMAX, and integrates them with deep-learning models, LSTM or CNN. In this approach, the residuals from the statistical model are used as inputs for the deep learning models, allowing the hybrid model to capture remaining patterns and nuances via deep learning. For Bitcoin and Ethereum price prediction in study 2, using LSTM as a base learner improved model accuracy. The results confirm that LSTMs hold an advantage as base learners because they are better equipped to handle the temporal and long-term dependencies of non-stationary data [21, 22]. This study advances existing work by demonstrating that LSTM, as a base learner, can outperform other modeling architectures even with limited input data. This is evidenced by the fact that CNN-based hybrid models (CNN+ARIMA/ARIMAX) do not achieve the same accuracies as an LSTM hybrid model, even with a full set of data (MI, TS, TC, and GT).

It is also worth noting the stability of hybrid models compared to both ensemble and multi-modal models. The range in Z-scores for multi-modal models is equivalent to 1 standard deviation, while the range in Z-scores for hybrid models is 0.32 for Bitcoin and 0.05 for Ethereum. The significant difference in performance between the two multi-modal models suggests that these models are sensitive to the type of data accepted by each base learner.

6.2 Feature Importance and Temporal Dynamics

The empirical results provide clear insights into which features contribute most significantly to cryptocurrency price prediction. Market indicators consistently emerged as the strongest predictors across both studies, as evidenced by the superior performance of models trained solely on market data. This is particularly apparent in Study 1, where the LSTM model using only market indicators achieved the lowest RMSE (4.56e-03) and MAPE (2.10e-05) scores (Table 4), outperforming more complex models that incorporated social media features. The relative importance of different features can be observed through the performance degradation when adding social media signals. For instance, in Study 2 (Table 5), the LSTM+ARIMA hybrid model using only market indicators achieved an RMSE of 1725.3, while the addition of social media features (Twitter, sentiment, Google Trends) resulted in slightly higher RMSE values (1866.9-1921.1). This suggests that while social media signals contain predictive value, their contribution is often overshadowed by noise, at least within a daily prediction horizon.

The temporal aggregation of social media data proved to be a critical factor in model performance. Study 1 aggregated Twitter data hourly, while Study 2 used daily aggregation for Twitter, sentiment, and Google Trends data. Neither approach demonstrated significant improvement in prediction accuracy, despite previous research showing strong correlations between Twitter volume and cryptocurrency prices. The impact of different lag periods on prediction accuracy was particularly noteworthy. Critien et al. [23] found that a 3-day lag outperformed a 7-day lag, with suggestions that a 1-day lag might be optimal due to temporal proximity, our results indicate that even a 1-day lag failed to overcome the inherent noise in social media signals. Further evidence of complex temporal dynamics comes from other studies. Said et al. observed increased volatility on the second day with 5.0% significance, while Kjærland et al. [24] found that Google search impacts were more than twice as strong with two lags (0.42% price increase) compared to a single lag (0.16%). These findings suggest that the relationship between social media signals and price movements may operate on longer time scales than initially assumed.

The limitation to Twitter and Google Trends data may have also impacted our findings. Research by Mai et al. [25] suggests that cryptocurrency prices are significantly influenced by a 'silent majority' of users who are less active on social media. Their work indicates that Internet forum discussions may have stronger predictive power than tweets, particularly at the inter-day level. This suggests that future short-term prediction models might benefit from incorporating forum data to reduce the noise observed in Twitter and Google signals.

6.3 Dynamics between market indicators and social media data

The comparative performance of market-based and sentiment-based predictions reveals several important insights. Market indicators demonstrate remarkable predictive power, particularly for short-term forecasting. This is evidenced by the consistent performance of market indicator-only models across both Bitcoin and Ethereum predictions (Tables 4-6). Two of the three experimental setups clearly show that models trained exclusively on market indicators outperform those using the full feature set, suggesting that market variables are more reliable predictors of price movements.

There is also a notable difference in the range of Z-scores across the models built to predict prices for Bitcoin and Ethereum. Said et al. [26] establish in their study that Ethereum is more robust to shocks from social media (YouTube, news searches, twitter sentiment); models that predict the price for Ethereum display a more robust performance across feature combinations. Presented in Figure 8 is a density plot that estimates the density function of RMSE Z-scores.

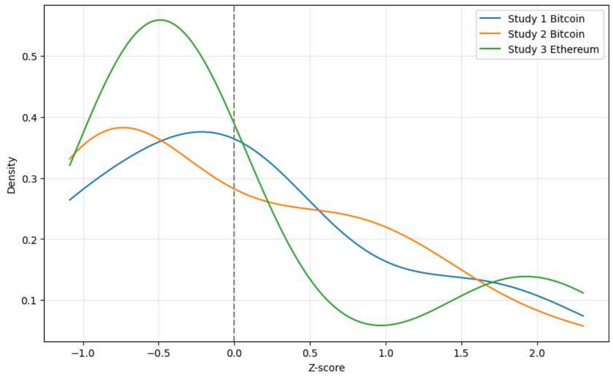

Fig. 8: Probability Distribution of Z-scores

Sentiment analysis, while intuitively appealing, showed limited effectiveness in improving short-term predictions. This is most clearly demonstrated in Study 2's results (Table 5), where models incorporating sentiment features often performed worse than their market-only counterparts. For instance, the addition of sentiment features to the CNN model increased the RMSE from 3679.9 to 4520.3. At the 1-hour prediction horizon, Twitter sentiment appeared to have an almost negligible influence on price, while market indicators maintained strong predictive power. The temporal horizon of predictions emerged as a crucial factor in model performance. While the 1-day prediction horizon showed improved accuracy compared to hourly predictions, the current approach of aggregating data into single records per time period (hourly or daily) may limit model generalizability. This limitation is particularly evident in the performance degradation when adding social media features, suggesting that the models struggle to effectively incorporate this additional information within the compressed temporal framework. Future work might benefit from larger datasets spanning longer periods to better capture the interplay between market indicators and sentiment signals across different time scales.

7 Conclusion

This paper compares the experimental results of two studies that aim for short-term price prediction of Bitcoin and Ethereum. The paper evaluates results over singluar models, different network architectures, and a variety of feature combinations. Both studies relied on market indicators and social media data. The aim of this comparitive study was to gain insights, and develop an understanding, of short-term price prediction models. The two studies make several notable contributions to the existing literature on cryptocurrency price prediction. First, they demonstrate the superior effectiveness of hybrid architectures, particularly those combining statistical methods (ARIMA/ARIMAX) with deep learning (LSTM). This is evidenced by the consistent outperformance of hybrid models across both studies, as shown in Tables 4 and 5, where hybrid models consistently achieved lower error rates and better Z-scores than singular models.

Second, the research provides empirical evidence challenging the assumed importance of social media sentiment in short-term price prediction. While previous studies have emphasized the role of social media signals, our results suggest a more nuanced relationship, particularly at shorter time horizons. This is clearly visualized in Figures 5-7, which show that models using primarily market indicators can achieve comparable or better prediction accuracy compared to those incorporating social media features.

Finally, the studies advance our understanding of model stability and reliability. The comparative analysis of Z-scores across different model architectures reveals that hybrid models achieve not only better accuracy but also more consistent performance across different market conditions. This is particularly evident in the Ethereum predictions, where hybrid models maintained stable performance across various feature combinations, as shown in Table 6. These findings suggest that while social media sentiment may influence cryptocurrency prices, its impact is more complex and potentially less direct than previously assumed. The superior performance of market indicators, especially when combined with sophisticated modeling architectures, indicates that technical analysis remains a more reliable approach for short-term price prediction. Future research might benefit from exploring longer prediction horizons and alternative sources of sentiment data to better capture the relationship between social media signals and price movements.

Declarations

Acknowledgements The authors would like to acknowledge several masters students that helped in the implementation and running of the models presented within this study.
Author Contributions Double blind policy - Authors contributions will be added after review
Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.
Conflict of Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Funding The authors declare that there was no funding received for the research associated with this study.

References

[1] L. Kristoufek, "What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis," *PLOS ONE*, vol. 10, 04 2015.

[2] A. Khedr, I. Arif, P. P V, M. El-Bannany, M. Alhashmi, and M. Sreedharan, "Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey," *Intelligent Systems in Accounting, Finance and Management*, vol. 28, 03 2021.

[3] Y. Hua, "Bitcoin price prediction using arima and lstm," *E3S Web of Conf.*, 2020.

[4] P. Mohanty, D. Patel, P. Patel, and S. Roy, "Predicting fluctuations in cryptocurrencies' price using users' comments and real-time prices," in *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2018, pp. 477–482.

[5] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering,"

Journal of Computational and Applied Mathematics, vol. 365, p. 112395, 2020.
[6] S. Ji, J. Kim, and H. Im, "A comparative study of bitcoin price prediction using deep learning," *Mathematics*, vol. 7, no. 10, 2019.
[7] X. Huang, W. Zhang, Y. Huang, X. Tang, M. Zhang, J. Surbiryala, V. Iosifidis, Z. Liu, and J. Zhang, "LSTM based sentiment analysis for cryptocurrency prediction," *CoRR*, vol. abs/2103.14804, 2021. [Online]. Available: https://arxiv.org/abs/2103.14804
[8] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. Muller, "Deep learning for time series classification: a review," *CoRR*, vol. abs/1809.04356, 2018. [Online]. Available: http://arxiv.org/abs/1809.04356
[9] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," *Entropy*, vol. 21, no. 6, 2019.
[10] J. Abraham, D. W. Higdon, J. Nelson, and J. Ibarra, in *Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:52950647
[11] O. Fathi, "Time series forecasting using a hybrid arima and lstm model," *Velvet consulting*, pp. 1–7, 2019.
[12] A. Tanwar and V. Kumar, "Prediction of cryptocurrency prices using transformers and long short term neural networks," in *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, 2022, pp. 1–4.
[13] K. Murray, A. Rossi, D. Carraro, and A. Visentin, "On forecasting cryptocurrency prices: A comparison of machine learning, deep learning, and ensembles," *Forecasting*, vol. 5, no. 1, pp. 196–209, 2023. [Online]. Available: https://www.mdpi.com/2571-9394/5/1/10
[14] J. Wu, X. Zhang, F. Huang, H. Zhou, and R. Chandra, "Review of deep learning models for crypto price prediction: implementation and evaluation," 2024.
[15] T. C. Nauen, S. Palacio, F. Raue, and A. Dengel, "Which transformer to favor: A comparative analysis of efficiency in vision transformers," 2024. [Online]. Available: https://arxiv.org/abs/2308.09372
[16] K. Wołk, "Advanced social media sentiment analysis for short-term cryptocurrency price prediction," *Expert Systems*, vol. 37, 11 2019.
[17] A. N. Youssfi, M. Bouchakwa, and Y. Jamoussi, "Bitcoin price prediction considering sentiment analysis on twitter and google news," in *Proceedings of the 27th International Database Engineered Applications Symposium*, ser. IDEAS '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 71–78. [Online]. Available: https://doi.org/10.1145/3589462.3589494
[18] G. Dutta, "Bitcoin tweets - 16m tweets with sentiment tagged," Aug 2021. [Online]. Available: https://www.kaggle.com/datasets/gauravduttakiit/bitcoin-tweets-16m-tweets-with-sentiment-tagged
[19] Yahoo!Finance, "Crypto real time prices and latest news," 2022, updated Daily. [Online]. Available: https://finance.yahoo.com/crypto/
[20] A. Altan, S. Karasu, and S. Bekiros, "Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques," *Chaos, Solitons & Fractals*, vol. 126, pp. 325–336, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960077919302632
[21] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting cryptocurrency prices using lstm, gru, and bi-directional lstm: A deep learning approach," *Fractal and Fractional*, vol. 7, no. 2, 2023. [Online]. Available: https://www.mdpi.com/2504-3110/7/2/203
[22] N. Patel, N. Vasani, N. K. Jadav, R. Gupta, S. Tanwar, Z. Polkowski, F. Alqahtani, and A. Gafar, "¡i¿f-lstm¡/i¿: Federated learning-based lstm framework for cryptocurrency price prediction," *Electronic Research Archive*, vol. 31, no. 10, pp. 6525–6551, 2023. [Online]. Available: https://www.aimspress.com/article/doi/10.3934/era.2023330
[23] J. V. Critien, A. Gatt, and J. Ellul, "Bitcoin price change and trend prediction through twitter sentiment and data volume," *Journal of Financial Innovation*, vol. 8, 05 2022.
[24] F. Kjærland, M. Meland, A. Oust, and V. Øyen, "How can bitcoin price fluctuations be explained?" *International Journal*

of Economics and Financial Issues, vol. 8, no. 3, p. 323–332, May 2018. [Online]. Available: https://www.econjournals.com/index.php/ijefi/article/view/6446

[25] F. Mai, J. Shan, Q. Bai, S. Wang, and R. Chiang, "How does social media impact bitcoin value? a test of the silent majority hypothesis," *Journal of Management Information Systems*, vol. 35, pp. 19–52, 01 2018.

[26] F. F. Said, R. S. Somasuntharam, M. R. Yaakub, and T. Sarmidi, "Impact of Google searches and social media on digital assets' volatility," *Humanities and Social Sciences Communications*, vol. 10, no. 1, pp. 1–17, December 2023.