Int. J. Advance Soft Compu. Appl, Vol. 18, No. 1, March 2026
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Bitcoin Price Forecasting Leveraging X Data and Sentiment Indicators Via an LSTM-Enhanced Deep Learning Architecture

Yunus Ozen and Mohammed Amen Azal Alwindawi*

Department of Computer Engineering, Yalova University, Yalova, Turkey
*Correspondence: 195105012@ogrenci.yalova.edu.tr

Abstract

Role of sentiment analysis in forecasting cryptocurrency market trends. The research employs a Long Short-Term Memory (LSTM) deep learning model integrated with sentiment data to enhance predictive accuracy. Tweets related to Bitcoin were collected during 2022–2023 and analyzed using the VADER tool, which classified them into positive, neutral, and negative categories. Sentiment scores were combined with historical Bitcoin prices to investigate their correlation before developing predictive LSTM models. The findings reveal that social media sentiment significantly influences Bitcoin price fluctuations, particularly during periods of high volatility. Incorporating sentiment in-formation improved the model's performance, as indicated by lower error metrics (MAE, RMSE, and MSE) compared to price-only models. This demonstrates that market emotions expressed on social media can serve as valuable predictive indicators in cryptocurrency forecasting. The study concludes that integrating sentiment data enhances the accuracy of Bitcoin price prediction and highlights the importance of emotional factors in market dynamics. Future research should extend this approach to other cryptocurrencies, multilingual sentiment contexts, and real-time analysis to further advance forecasting capabilities in digital financial markets.

Keywords - cryptocurrency; deep learning; LSTM; sentiment analysis; VADER.

1 Introduction

The utilization of artificial intelligence (AI) and machine learning (ML) in financial markets has changed the way investment strategies are developed, particularly in the volatile and high-risk cryptocurrency field. One of the innovative AI ideas, sentiment analysis that helps collect and analyze the mood of people from textual data containing the tweets, news, articles and forums, improves the accuracy of the estimated prices and potential trading signals [1-5]. In cryptocurrency markets, where investors are the biggest driver of price dynamics, using sentiment analysis coupled with ML algorithms has displayed the ability to predict short term trends and fluctuations in price. LSTM and Recurrent Convolutional Neural Networks (RCNN) are useful for understanding the temporal and semantic features of data that will help to solve the problem of the financial time series data [6-9]. This paper provides an extensive evaluation of AI models' execution in the cryptocurrency forecast and how social media

Received 5 December 2025; Accepted 20 February 2026

activity, news sentiment, affect the model's efficiency in aiding investor decisions. By these viewpoints, this chapter demonstrates the potential of the sentiment analysis approach to become a predictive tool in the competitive trading environment and provides a basis for further investigation into the combination of AI with sentiment for improving financial prognosis and algorithmic trading [10-14]. Unlike prior studies that primarily emphasize complex hybrid architectures, this study focuses on a transparent and interpretable LSTM-based framework evaluated on real-world Bitcoin data collected during the 2022–2023 bear market period. A distinctive contribution of this work is the explicit quantification of the temporal lag effect of social media sentiment, revealing that sentiment exerts its strongest influence on Bitcoin price changes with a one-day delay. This provides practical insight into short-term market behavior under high volatility conditions.

1.1. Problem Statement

It is as a result challenging to make the right forecast because the markets in the financial sector are volatile and competitive. Conventional approaches to forecasting such as time series, econometric models are less effective for dynamic market conditions and short-term fluctuations that are typical in these markets. In addition, even though there are references about utilizing CNNs and LSTMs in the financial data, the problem of combining different sources, including historical prices, the sentiment of social media posts, and the overall market conditions, remains challenging. This study aims at solving these problems by trying to establish the utilization of deep learning models, especially CNNs and LSTMs for the prediction of the financial market's future performance, specifically the cryptocurrency and stock market.

1.2. Research Objectives

This study's key aims are as follows:

- LSTMs hybrid combinations to determine the extent to which these deep learning models are capable of enhancing the precision of the financial market forecasts.
- To quantify the competencies of deep learning models with historical prices, on-chain data, trading volumes, and social media sentiment as input data;
- For the purpose of comparing the results and effectiveness of ensemble learning techniques and feature selection methods in increasing model stability and de-creasing overfitting (e.g. Light GBM, AdaBoost, Boruta);
- To understanding the best approaches of reducing the black box problem of deep learning models that can be applied in financial forecasting;
- To suggest an architecture for deep learning prediction of financial data that lays stress on the three components of predictive accuracy, computational tractability, and interpretability.

1.3. Research Contribution

This paper brings significant achievements to the area of financial forecasting using deep learning:

- It shows that the use of social media sentiment and other non-conventional data sources in improving the deep learning models in volatile markets;
- The work also introduces the concepts of ensemble learning as well as feature se-lection to enhance performance, flexibility as well as reliability of the models especially where the financial markets' data is normally complex and full of noise;
- It outlines a strategy for constructing better and less rigid financial forecasting models that can be used by financial practitioners and researchers to make better decisions in volatile markets.

This paper delivers an un-depth analysis of AI models' execution in the crypto-currency forecast and how social media activity, news sentiment, affect the model's efficiency in aiding investor decisions. By these viewpoints, this chapter demonstrates the potential of the sentiment analysis approach to become a predictive tool in the competitive trading environment.

In [15], the authors investigate at utilizing sentiment analysis with empirical mode decomposition (EMD) and deep learning models to improve the ability to predict the price of Bitcoin. The aim of analysis is to assess the impact of public opinion especially from X "formerly Twitter" on the price of Bitcoin. The methodology involves a dual-network model: the first network uses EMD to analyze historical prices into IMFs then the obtained IMFs undergo filtering through an LSTM network to remove noise from fluctuations in prices. The resultants of both the networks are summed up to determine final prediction values.

While in [16], the authors compare the applicability of new generation machine learning approaches for determining the value of several selected cryptocurrencies including Bitcoin, Ethereum, Ripple, Litecoin, and others. the methodology consists in training of such models on the historical data on the cryptocurrency prices and in their division according to the periods occurred before and after the beginning of the Covid-19 pandemic to consider the effects of such shifts on the quality of the models. Quantitative benchmarking metrics that are commonly applied are MAPE, MAE, RMSE and correlation coefficients together with investor-oriented metrics like MDA and return scores. Finding indicates that ensemble learner or Lights GBM outcompete others in anticipating Bitcoin and Ethereum prices with high accuracy and Simple RNN for Ripple. It is seen from the study that after the Covid-19 shock the volatility has enhanced the model ac-curacy.

Article [17] aimed at the optimization of Bitcoin trading by applying deep learning models for the prediction of price trends instead of the price level. The main goals are to compare the performance of different model structures such as CNN-LSTM, LSTNet, TCN and ARIMA (baseline model) and to assess the profitability of trading approaches realized from these predictions. A range of feature-selection approaches, including Boruta, GA, and Light GBM, is used for improving the number of input data dimensions. The models are built to identify the directional movement of the prices, and three trading strategies are evaluated for their profitability based on the on-chain data. The findings also established that the CNN-LSTM model combined with Boruta feature selection was most ac-curate with 82.44%, better than the benchmark model, namely the ARIMA model. Regarding the trading strategies, only the long-and-short strategy based on the CNN-LSTM yielded annualized return as high as 66.54%.

In [18] authors seek to increase the reliability of predicting Bitcoin's prices by using Deep Q-Network (DQN) model. The approach includes using a large amount of data from these features and training a DQN with a reward function consisting of a set of factors such as confidence scaling, punishment for continuous errors, and time-based adjustments for recent trends. Such a reward structure is meant to enhance the flexibility of the model used and its accuracy in capturing market characteristics. Performance evaluation likewise indicates that the model has a high F1-score of 95% that indicates high level of possibility for trading.

Compared to prior studies, this study introduces numerous unique additions to the field of cryptocurrency forecasting. While many works (e.g., Arslan [15], Bouteska et al. [16], and Muminov et al. [18]) focus on improving model accuracy through complex hybrid architectures like CNN-LSTM, EMD, or reinforcement learning, this study emphasizes a more interpretable LSTM-based model that balances predictive performance with practical applicability. Notably, it uniquely analyzes the temporal lag effect of sentiment on Bitcoin price changes, identifying a peak influence at a one-day delay, which has not been explicitly quantified in most related literature. Furthermore, the dataset used in this study—76,797 tweets spanning 2022 to 2023—offers a real-world and timely reflection of public opinion during actual Bitcoin market cycles. Detailed text preprocessing steps, including bot filtering and noise reduction, ensure high data quality and model reliability. Unlike many prior studies, which treat sentiment as a static input, this work highlights how sentiment's predictive power is amplified during periods of market volatility, offering new behavioral insights into trader psychology. Finally, this research contributes to addressing the deep learning "black-box" problem by prioritizing model interpretability, making it more suitable for practical financial decision-making.

2 Methodology

In modeling, the quality of the data is very crucial in that it is the data that is used in the development of the bulk of the other models. This chapter is devoted to the description of the dataset selection and data preprocessing performed before sentiment analysis and before studying the correlation between sentiment and Bitcoin price changes. The dataset for this study was obtained from Mendeley [19] and comprises 76,797 tweets posted between 01/01/2022 and 22/06/2023. All the tweets are about Bitcoin which is a digital currency that has been widely discussed and which has high fluctuation rate.

2.1. Model Block Diagram

The part of the diagram in Figure 1 demonstrates the training and evaluation process for sentiment-driven price prediction, using both sentiment analysis and historical datasets. The pipeline begins with loading the sentiment data, followed by data cleaning and text cleaning. Afterward, a sentiment extraction model is applied to acquire sentiment scores. After that, the historical data is loaded and combined with the processed sentiment data.

The merged dataset is then separated into training and testing sets. The training set is put into the LSTM model for training, while the test set is utilized for assessment purposes. The way the model performs is evaluated with metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). This systematic technique assures strong preprocessing, model training, and evaluation, resulting in accurate and dependable predictions. Figure 1 illustrates the overall workflow of the proposed sentiment-driven Bitcoin price forecasting model.

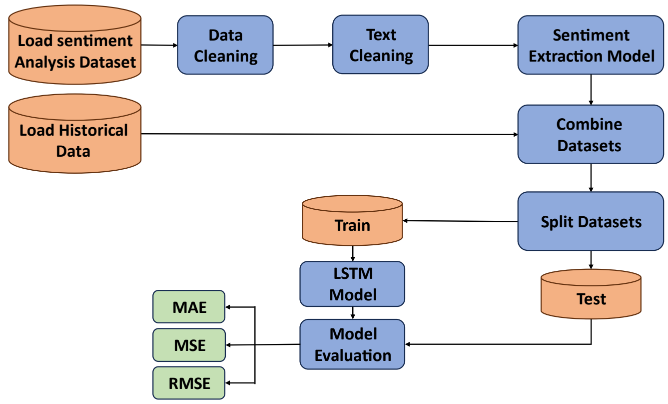

Figure 1. Model block diagram.

2.2. Dataset Selection and Preprocessing

The dataset extracted from Mendeley involving 76,797 Tweets which had been re-leased between January 1, 2022, and June 22, 2023. This period reflects the bitcoin performance on a normal bitcoin cycle on a bearish wave. As this dataset is devoted to the materials connected with Bitcoin, it will be quite suitable for the examination of the public opinion on the subject of Bitcoin and its price indices.

Based on the use of the proposed model, it is the tweet text that is important for performing sentiment analysis because it captures people's position and emotional attitude to Bitcoin. This dataset is particularly useful to evaluate changes in the public sentiment and to look how these changes may correlate to changes in the price of Bitcoin. Another advantage of this dataset is that sentiments of the tweets are already labeled and scored.

2.3. Data Formatting

To provide for uniformity and to make the further processing unhampered, data formatting was conducted for all fields of the dataset. For instance, the date field was normalized into a single date format for performing time-based analysis. Also, we made sure that all the text data were of equal encoding format, UTF-8, to avoid problems when cleaning the text and analyzing its sentiment. Table 1 shows data formatting. Table 1 summarizes the data formatting procedures applied in this study.

Table 1. Data formatting.

Action	Description	Purpose
Standardization of Date	Convert all dates to a consistent format (e.g., YYYY-MM-DD).	Ensures consistency for temporal analysis.
Uniform Encoding	Convert all text data to UTF-8 encoding.	Prevents encoding issues in text processing.

Format Adjustment	Adjust categorical fields for consistency (e.g., yes/no or 0/1 for binary fields).	Facilitates smoother analysis and modeling.

2.4. Data Cleaning

Data cleaning is a vital process in the data preprocessing step for dealing with the dataset. The idea was to filter out any kind of noise, additional entries or paradoxes that may distort the result or introduce bias. Table 2 shows the data cleaning stages. Table 2 presents the main data cleaning stages performed on the dataset.

Table 2. Data cleaning stages.

Action	Description	Reason
Removal of Null Rows	Remove rows with missing or null values in any field.	To prevent missing data from skewing the analysis.
Removal of Duplicates	Remove duplicate entries based on tweet text and date.	Ensures each tweet is counted only once.
Filter Unwanted Words	Eliminate irrelevant terms such as "free," "earn bitcoin," "make money," etc.	Reduces noise from spam or promotional content.
Remove Inconsistent Data	Remove any records that are obviously irrelevant or erroneous (e.g., bots).	Improves the quality and relevance of data.

2.5. Preliminary Analysis

Once the dataset had been cleaned, initial analysis was conducted to allow a pre-liminary view on the periodicity of the tweet distribution.

This preliminary analysis proved to be helpful in understanding how active the conversation regarding Bitcoin was throughout the chosen timeframe and also in finding out time intervals where sentiment change is likely to be correlated with significant events in Bitcoin market. Table 3 shows the preliminary analysis.

Table 3. Preliminary analysis of tweet distribution

Action	Description	Purpose
Count Tweets by Month	Count the number of tweets per month.	Identify trends and periods of high tweet volume.
Identify Patterns	Identify months with significant differences in tweet volume (e.g., spikes or dips).	Recognize events that could influence sentiment.
Observation of Trends	Analyze the tweet volume differences between the first 6 months and the rest.	Highlight potential events impacting Bitcoin sentiment.

2.6. Text Cleaning

Before making the text of the tweet suitable for sentiment analysis, the text was cleaned to eliminate unnecessary and unwanted texts that could distort sentiment analysis. The steps in Table 4 shows implemented as part of this text cleaning process.

Table 4. Implemented as part of this text cleaning process.

Action	Description	Purpose
Remove Links	Eliminate URLs or links from tweets.	To focus analysis on tweet content, not external resources.
Remove Numbers	Remove any numerical values that do not contribute to sentiment analysis.	To avoid bias from specific numeric data (e.g., Bitcoin prices).
Remove Special Characters	Eliminate special characters, punctuation marks, and unnecessary symbols.	Reduce noise and irrelevant data from the text.
Remove Stopwords	Eliminate common stopwords (e.g., "the," "and," "is") that don't carry meaningful sentiment.	Focus on words that contribute to sentiment.

2.7. Dataset Visualization

The analysis of the dataset can be done using a plethora of data science tools and techniques. This visualization is made of several sections each of which focuses on different aspects of the dataset. The analysis includes data over time, number of tweets throughout periods of various time, and the distribution of sentiment scores.

2.7.1. Number of Crypto Tweets Over Time

The number of tweets containing crypto on a timeline from January 2022 to July 2023. The horizontal axis stands for the date, the vertical axis (in a logarithmic form) is the number of tweets. Figure 2 shows the number of cryptocurrency-related tweets over time from January 2022 to July 2023.

The data implies that the frequency of tweets has remained constant over the du-ration covered as evident by the green bars approximated to be of equal height. It is possible to also note that there are differences between tweet counts, nonetheless these are not very significant and might be caused by daily or seasonal activity.

2.7.2. Distribution of Tweets by Year

The number of the tweets done in a year, the difference being between 2022 and 2023. In 2022, the tweets recorded where 45,367 while in 2023, the tweets recorded where only 31,341. This means that the percentage of tweets has reduced from one year to the next, that is, the percentage of tweets has reduced. There are several reasons for this decrease, including: different platform policies, user behavior, and temporal or spatial Tweeting trends, or external circumstances that resulted in a general Tweeting volume trend. The chart has proper axis title and special labels to explain the data so it is relaxed to decipher the chart. Figure 3 presents the yearly distribution of tweets across 2022 and 2023.

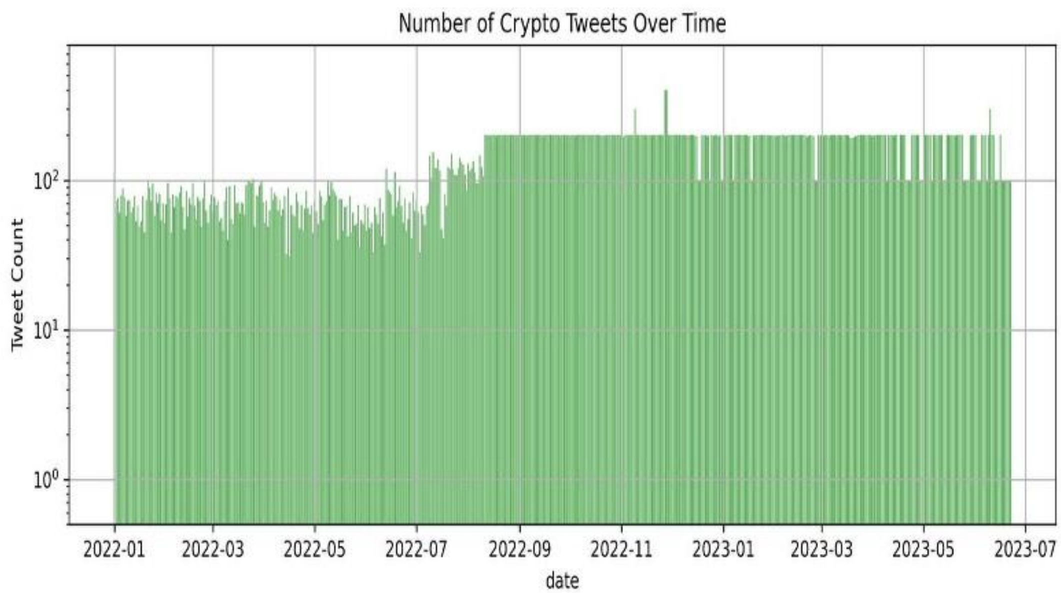

Figure 2. Number of crypto tweets over time.

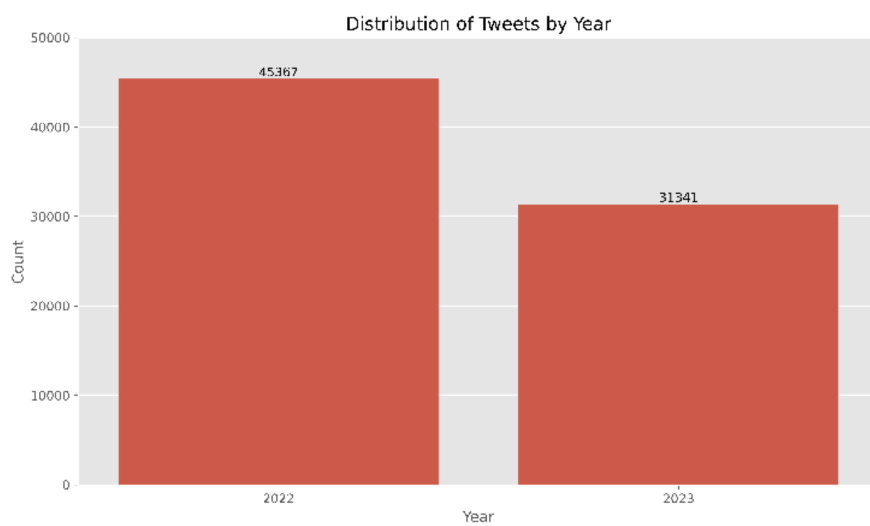

Figure 3. Distribution of tweets by year.

2.7.3. Distribution of Tweets by Month

The monthly distribution of tweets involving this hashtag for 2022 and 2023 separately to show the difference in patterns of the two years. In 2022 the tweet activity was lower in the first half of the year with the minimum monthly count of 1,937 in April and the maximum of 2,994 in July. But the activity was highest in August with 2608 and the highest number of tweets published in November with 6490 and slightly reduced activity in December with 5786. However, the year 2023 started with higher activity level; making 5889 tweets in January and 6162 in March which was highly active month of the year. Figure 4 depicts the monthly distribution of Bitcoin-related tweets for both years.

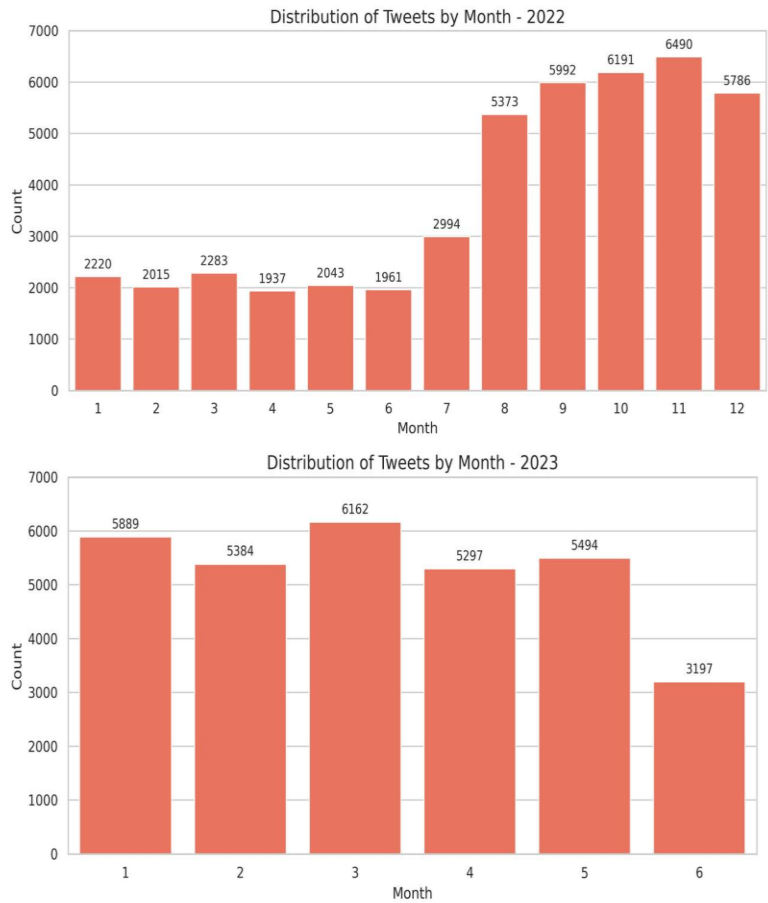

Figure 4. Distribution of tweets by month.

2.7.4. Sentiment Score Distribution by Label

The chart is the breakdown of the sentiment scores by the positive, neutral and negative labels determined by the VADER sentiment analysis tool. Most of the data is clustered around a sentiment score of zero and the blue line of the neutral sentiments is much more pronounced. The majority of the positive sentiment scores is shown in red and ranges mostly above 0 with a density at approximately 0.25. Sentiment scores below 0 that depicts the negative sentiment is located in the region below the x-axis as seen in purple color, and the peaks are relatively lower than the other two categories. The above chart shows that, the majority of the tweets are often categorized as objective and they were fewer that fall under the positive and negative sentiments. This distribution indicates that the analyzed content is primarily balanced, and there is a smaller share of highly positive and negative emotions.

Figure 5 illustrates the distribution of sentiment scores across positive, neutral, and negative classes.

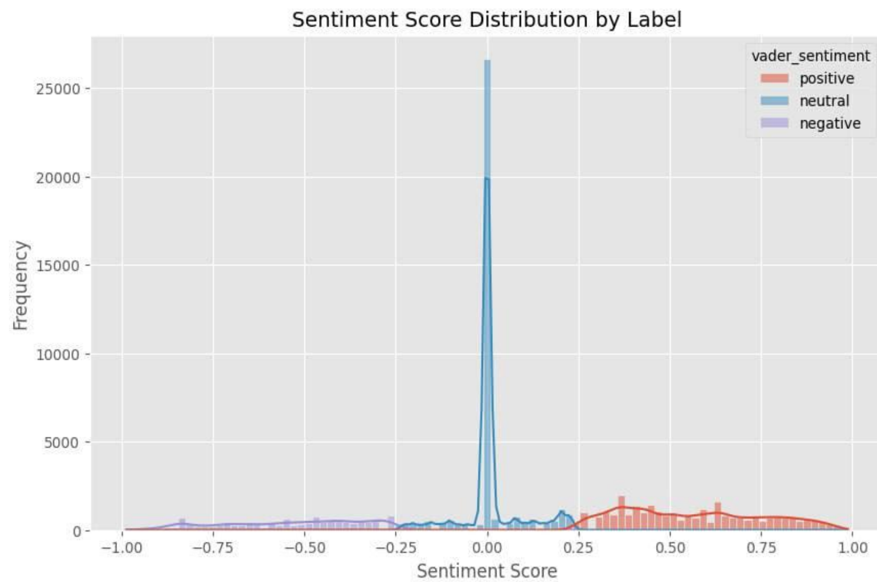

Figure 5. Sentiment score distribution by label.

2.8. Sentiment Analysis

BERTweet is built to capture the nature of tweets including abbreviations, hashtags, and emojis making it suitable for this dataset. It provides sentiment labels and scores that category each tweet as positive, negative or neutral in terms of the expressive text.

Nevertheless, we noted a problem with the results from BERTweet since its sentiment analysis leaned much more towards negative polarity. We hypothesize that this bias stemmed from the way in which BERTweet had been fine-tuned on its training data, causing it to perform better when it is tasked with detecting negative sentiment than positive sentiment.

In order to solve this problem, we transitioned to using VADER Sentiment Analysis. VADER generates polarity scores on a scale of negative one for the most negative to positive one for the most positive tweets and the type of tweets is also categorized as positive, neutral and negative.

2.9. Historical Data

Next to the obtained sentiment scores, we included historical daily Bitcoin prices retrieved from Kaggle [20] we matched the sentiment scores with historical Bitcoin prices in order to analyze how sentiment impacted or associated with Bitcoin prices.

We might be able to investigate how changes in public sentiment for better or for worse affected the stock and its fluctuations. Table 5 shows the historical data structure.

Table 5. Historical data structure.

Action	Description	Purpose
Kaggle Cryptocurrency Data	Daily historical price data for Bitcoin (and other cryptocurrencies).	To analyze the relationship between sentiment and Bitcoin price movements.
Date Range	Aligns with the sentiment data from January 1, 2022, to June 22, 2023.	Ensures compatibility with sentiment analysis timeline.

2.10. Sentiment Data can be Compared with Price Data in The History

The last process that was followed involved combining the sentiment scores with the historical price data obtained. This integration was achieved by aggregating the senti-ment data to create two new fields:

- Daily Mean Sentiment Score: This field gives the average of the sentiment score for that particular day giving an outlook of the general public trend in a particular day regarding the Bitcoin as extracted from the tweets;

- Daily Tweet Count: This field contains an estimate of total tweets on a particular day, it will help the user to estimate the total discussion on Bitcoin on that specific day;

- We integrated both sentiment fields with the daily Bitcoin price data and got a set of data that gives an indication of sentiment on specific days together with the price. Table 6 shows the full dataset characteristics.

Table 6. Full dataset characteristics.

Field	Description
Date	The date of the tweet or the corresponding daily price.
Tweet Text	The content of the tweet after text cleaning.
Sentiment Score	The polarity score (ranging from -1 to +1) indicating tweet sentiment.
Sentiment Label	The classification of the tweet as positive, neutral, or negative.
Daily Mean Sentiment Score	The average sentiment score for all tweets on a given day.
Daily Tweet Count	The total number of tweets posted on a given day.
Bitcoin Closing Price	The closing price of Bitcoin on the corresponding date.

2.11. Building LSTM Model

In this study, a deep learning model utilizing Long Short-Term Memory (LSTM) was built to forecast Bitcoin price movements using sentiment ratings collected from X-data. The model design comes with two LSTM layers, followed by a dense output layer.

The selected LSTM architecture represents a deliberate trade-off between model complexity and interpretability. Preliminary experiments with deeper architectures and alternative hyperparameters did not yield significant performance improvements. Therefore, a two-layer

LSTM configuration was adopted to reduce overfitting risk and ensure stable convergence, while maintaining sufficient capacity to capture temporal dependencies in the data.

The first LSTM layer includes 256 hidden units and is configured to return sequences to the subsequent LSTM layer. The second LSTM layer comprises 128 units and outputs a single hidden state to a Dense layer with a linear activation function, suitable for regression tasks. The input shape for the model is defined as (window_size, num_features), where window_size = 7 corresponds to a 7-day time window of historical data. The model is trained utilizing the Mean Squared Error (MSE) loss function and optimized with the Adam optimizer at a learning rate of 0.001 percent. The training was done across 100 epochs with a batch size of ten, and the results were validated using a separate validation set. This LSTM architecture was chosen for its ability to model temporal dependencies in time series data while addressing the vanishing gradient problem. Commonly found in standard RNNs. Figure 6 shows the architecture of the proposed LSTM model.

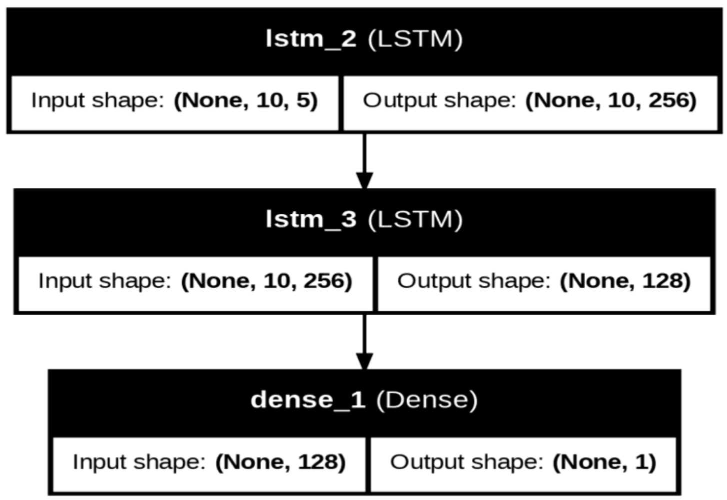

Figure 6. LSTM model.

An early stopping mechanism was previously set (even if commented out) and thus it is not used in the current form to stop the model at, for instance, epoch 18 due to overfitting and diminishing returns on validation set accuracy. These settings are used to control the batch size, and to monitor the validation, but as there is no early stopping this will take longer or produce inferior results if the model starts too overfit.

2.12. LSTM Training Parameters

The training process was carried out using a sequential deep learning model con-figured with standard parameters. The dataset was divided into batches of size 10, and the model was trained for a maximum of 100 epochs with a fixed sample order across iterations (i.e., no shuffling applied). A portion of the dataset was reserved for validation purposes to assess model generalization. While early stopping was considered as a regularization mechanism, it was not ultimately applied during final training. This configuration ensured consistent training dynamics and facilitated the evaluation of performance across epochs. Table 7 shows training parameters. Table 7 lists the training parameters used for the LSTM model.

Table 7. Training parameters.

Parameter	Value
Batch size	10
Epochs	100
Shuffle	FALSE
Validation data	(x_val, y_val)
Early stopping	Not used (commented out)
Restore best weights	Not used (commented out)

3. Results and Discussion

This work attempted to evaluate the correlation between the sentiments on the social media and Bitcoin value with an aim of determining if sentiments could be used to predict the market direction. In this study, to explore the temporal pattern in the relationship between sentiment and price change, the authors used sentiment analysis tools, and historical prices.

The LSTM model was trained using the combined dataset of sentiment scores and historical Bitcoin prices. Training and validation losses decreased over 100 epochs, showing good model learning without severe overfitting. Figure 7 shown the model achieved low training loss (~0.0024) and validation loss (~0.00057).

Table 8 presents the performance results of a predictive model with and without sentiment data. The goal here is to evaluate whether including sentiment improves the prediction of Bitcoin prices. Including sentiment in the model improved predictive performance. MAE dropped from 0.0349 to 0.029, MSE from 0.00166 to 0.00135, and RMSE from 0.04077 to 0.0368. These improvements confirm the value of integrating sentiment data. Figure 7 presents the training and validation loss curves of the LSTM model.

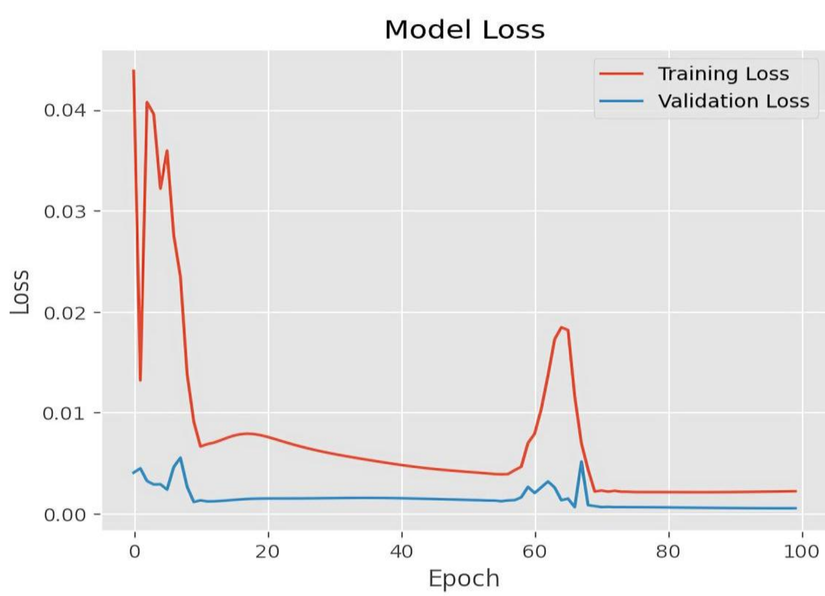

Figure 7. LSTM training history

Table 8. Sentiment predictions results.

Metric	No Sentiment Predictions	Sentiment Predictions
Mean Absolute Error (MAE)	0.0349	0.029
Mean Squared Error (MSE)	0.00166	0.00135
Root Mean Squared Error (RMSE)	0.04077	0.0368

Correlation analysis showed that Bitcoin price is most sensitive to sentiment changes with a 1-day lag (correlation = -0.06). Immediate reaction (lag 0) was weaker (0.02), and effects diminished further at lag 2 (0) and lag 3 (0.02). This suggests a short-term influence of sentiment on price movements. Table 9 shows the correlation between sentiment scores and Bitcoin price changes across different time lags (0 to 3 days). The aim is to examine whether public sentiment influences price movements immediately or with a delay.

The findings support the hypothesis that social media sentiment impacts crypto-currency markets, particularly in the short term. This has implications for using senti-ment-driven models in high-volatility trading environments.

Table 9. Correlation between sentiment scores and bitcoin price changes across time lags.

Lag (Days)	Correlation Coefficient	Observation
0	0.02	Weak positive correlation; limited immediate reaction.
1	-0.06	Highest correlation; negative, indicating significant response.
2	0	Neutral correlation; minimal sentiment influence.
3	0.02	Weak positive correlation; slightly delayed sentiment response.

The observed negative correlation at a one-day lag may reflect behavioral trading patterns commonly seen in cryptocurrency markets. Positive sentiment spikes can lead to short-term price increases, followed by profit-taking behavior on the subsequent day, resulting in price corrections. This phenomenon aligns with the well-known “buy the rumor, sell the news” effect, where investor optimism precedes corrective movements rather than sustained upward trends.

5. Conclusion and Future Work

This study intended to analyze the correlation between social media sentiment and Bitcoin value, with the goal of understanding whether sentiments may be utilized to anticipate market direction. In this study, to explore the temporal pattern in the correlation between sentiment and price change, the authors used sentiment analysis tools, and historical prices. The findings indicate that the price of Bitcoin is particularly reactive to one-day mood swings at -0.06. The prediction models that included sentiment data performed better, as seen by a decrease in Mean Absolute Error (MAE) and Root Mean Squared Error. On a specific level, the MAE for sentiment prediction was dropped from 0.0349 for the baseline to 0.029, but the RMSE went up from 0.04077 to 0.0368. When sentiment was factored into forecasts, the Mean Squared Error (MSE) dropped from 0.00166 to 0.00135. The study adds to the knowledge about how certain psychological parameters play out in the cryptocurrency markets and it deliberates their forecasting implications.

Beyond improving prediction accuracy, this study contributes practical insights into the timing of sentiment influence on Bitcoin markets. The identification of a short-term lag structure enhances the applicability of sentiment-aware forecasting models for traders and analysts operating in high-volatility environments, while preserving model transparency and interpretability.

In future work, we can extend this work by incorporating multilingual sentiment analysis to capture broader global market opinions. Additionally, experimenting with alternative deep learning architectures and advanced hyperparameter optimization techniques may further enhance predictive accuracy. Integrating real-time sentiment streams and on-chain metrics could also provide a more comprehensive understanding of cryptocurrency market dynamics.

References

[1] Wen, N. S.; Ling, L. S. Evaluation of cryptocurrency price prediction using LSTM and CNNs models. JOIV: Int. J. Inform. Visualization 2023, 7, 2016-2024. https://doi.org/10.30630/joiv.7.3-2.2344

[2] John, D. L.; Binnewies, S.; Stantic, B. Cryptocurrency price prediction algorithms: A survey and future directions. Forecasting 2024, 6, 637-671. https://doi.org/10.3390/forecast6030034

[3] Nguyen, D. T. A.; Chan, K. C. Cryptocurrency trading: A systematic mapping study. Int. J. Inf. Manag. Data Insights 2024, 4, 100240. https://doi.org/10.1016/j.jjimei.2024.100240

[4] Akter, M. S.; Shahriar, H.; Rahman, M. A.; Rahman, M.; Cuzzocrea, A. Early prediction of cryptocurrency price decline: a deep learning approach. In 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2023.

[5] Idowu, E. Advancements in financial market predictions using machine learning techniques. 2024. https://doi.org/10.20944/preprints202407.1075.v1

[6] Abou Tanos, B.; Badr, G. Price delay and market efficiency of cryptocurrencies: The impact of liquidity and volatility during the COVID-19 pandemic. J. Risk Financial Manag. 2024, 17, 193. https://doi.org/10.3390/jrfm17050193

[7] Jing, N.; Wu, Z.; Wang, H. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. Expert Syst. Appl. 2021, 178, 115019. https://doi.org/10.1016/j.eswa.2021.115019

[8] Ortu, M.; Uras, N.; Conversano, C.; Bartolucci, S.; Destefanis, G. On technical trading and social media indicators for crypto-currency price classification through deep learning. Expert Syst. Appl. 2022, 198, 116804. https://doi.org/10.1016/j.eswa.2022.116804

[9] Brim, A.; Flann, N. S. Deep reinforcement learning stock market trading, utilizing a CNN with candlestick images. Plos One 2022, 17, e0263181. https://doi.org/10.1371/journal.pone.0263181

[10] Bhatt, S.; Ghazanfar, M.; Amirhosseini, M. Sentiment-driven cryptocurrency price prediction: A machine learning approach utilizing historical data and social media sentiment analysis. Machine Learn. Appl.: Int. J. 2023, 10, 1-15. https://doi.org/10.5121/mlaij.2023.10301

[11] Feizian, F.; Amiri, B. Cryptocurrency price prediction model based on sentiment analysis and social influence. IEEE Access 2023, 11, 142177-142195. https://doi.org/10.1109/ACCESS.2023.3342688

[12] Saha, V. Predicting future cryptocurrency prices using machine learning algorithms. J. Data Anal. Inf. Process. 2023, 11, 400-419.

[13] Likhitha, B. B.; Raj, C. A.; Islam, M. S. U. Unveiling ethereum's future: LSTM-based price prediction and a systematic block-chain analysis. BIO Web Conf. 2024, 86, 01117. https://doi.org/10.1051/bioconf/20248600002

[14] Nair, M.; Marie, M. I.; Abd-Elmegid, L. A. Prediction of cryptocurrency price using time series data and deep learning algorithms. Int. J. Adv. Comput. Sci. Appl. 2023, 14, 338-347.

[15] Arslan, S. Bitcoin price prediction using sentiment analysis and empirical mode decomposition. Comput. Econ. 2025, 65, 2227-2248. https://doi.org/10.1007/s10614-024-10588-3

[16] Bouteska, A.; Abedin, M. Z.; Hajek, P.; Yuan, K. Cryptocurrency price forecasting–a comparative analysis of ensemble learning and deep learning methods. Int. Rev. Financ. Anal. 2024, 92, 103055. https://doi.org/10.1016/j.irfa.2023.103055

[17] Omole, O.; Enke, D. Deep learning for Bitcoin price direction prediction: models and trading strategies empirically com-pared. Financ. Innov. 2024, 10, 117. https://doi.org/10.1186/s40854-024-00643-1

[18] Muminov, A.; Sattarov, O.; Na, D. Enhanced Bitcoin price direction forecasting with DQN. IEEE Access 2024, 12, 29093-29112. https://doi.org/10.1109/ACCESS.2024.3367719

[19] Partida, D. "Tweets", Mendeley Data, V1. 2023. https://doi.org/10.17632/x7yvshrnxy.1

[20] Kaggle. Crypto currencies daily prices. 2025. https://www.kaggle.com/datasets/svaningelgem/crypto-currencies-daily-prices