



# **Cryptocurrency Price Prediction Using Social Media Post and Comments**

**Peer Mohammed S M, Pranav K.P<sup>2</sup>, Dr. V.Sangeetha<sup>3</sup>**

Student, Department of Computer Science with Data Analytics, Dr N.G.P. Arts and Science Collage, Coimbatore, India<sup>1</sup>

Student, Department of Computer Science with Data Analytics, Dr N.G.P. Arts and Science Collage, Coimbatore, India<sup>2</sup>

Associatet Professor, Department of Computer Science with Data Analytics, Dr N.G.P. Arts and Science Collage,  
Coimbatore, India<sup>3</sup>

**ABSTRACT:** Cryptocurrency markets exhibit high volatility and are significantly influenced by public sentiment expressed through social media platforms and online comments. This paper presents a novel approach for cryptocurrency price prediction that integrates sentiment analysis of social media data with traditional market indicators. We collect and analyze data from Twitter, Reddit, and cryptocurrency forums to extract sentiment features using Natural Language Processing techniques. These sentiment features are combined with historical price data and technical indicators to train predictive models using Long Short-Term Memory networks and Gradient Boosting algorithms. Our experimental evaluation on Bitcoin, Ethereum, and Litecoin demonstrates that incorporating social media sentiment improves prediction accuracy by 15-22% compared to models using only historical price data. The proposed system achieves Mean Absolute Percentage Error of 3.2% for Bitcoin price prediction over 7-day horizons, outperforming baseline approaches. Results indicate that social media sentiment serves as a valuable leading indicator for cryptocurrency price movements, particularly during periods of high market volatility.

**KEYWORDS:** Cryptocurrency Prediction, Sentiment Analysis, Social Media Mining, Natural Language Processing, LSTM Networks, Machine Learning

## **I. INTRODUCTION**

Cryptocurrencies have emerged as a revolutionary financial asset class, with Bitcoin, Ethereum, and thousands of alternative coins attracting significant investor interest and trading volumes exceeding trillions of dollars. Unlike traditional financial markets, cryptocurrency prices exhibit extreme volatility influenced by diverse factors including regulatory announcements, technological developments, market manipulation, and importantly, public sentiment expressed on social media platforms.

Social media platforms such as Twitter, Reddit, Telegram, and specialized cryptocurrency forums serve as primary channels for information dissemination and community discussion. Influential figures, developers, traders, and enthusiasts regularly share opinions, analyses, and predictions that can significantly impact market sentiment and subsequent price movements. The correlation between social media activity and cryptocurrency prices has been observed in numerous market events, suggesting that sentiment analysis could provide valuable predictive signals.

Traditional cryptocurrency price prediction approaches primarily rely on technical analysis using historical price data, trading volumes, and market indicators. While these methods capture price patterns and trends, they fail to account for the sentiment-driven nature of cryptocurrency markets. Machine learning techniques, particularly deep learning models such as Long Short-Term Memory networks, have demonstrated promising results in time series forecasting. However, their performance can be significantly enhanced by incorporating additional features derived from social media sentiment.

This research addresses the challenge of cryptocurrency price prediction by developing a comprehensive framework that integrates sentiment analysis of social media comments with traditional market indicators. Our contributions include: (1) development of a multi-source social media data collection and preprocessing pipeline, (2) implementation of advanced sentiment analysis techniques using transformer-based models, (3) design of hybrid prediction models

combining LSTM networks with sentiment features, and (4) comprehensive evaluation demonstrating improved prediction accuracy across multiple cryptocurrencies and time horizons.

## II. RELATED WORK

Cryptocurrency price prediction has attracted considerable research attention, with approaches ranging from classical statistical methods to advanced deep learning techniques. Early studies employed traditional time series models including ARIMA and GARCH, which capture temporal dependencies and volatility patterns. However, these linear models struggle to capture the complex nonlinear dynamics characteristic of cryptocurrency markets.

Machine learning approaches have demonstrated superior performance for cryptocurrency forecasting. Support Vector Regression and Random Forest models have been successfully applied, though they require manual feature engineering and struggle with long-term dependencies. Deep learning architectures, particularly Recurrent Neural Networks and their variants such as LSTM and GRU networks, excel at modeling sequential data and capturing temporal patterns in price movements.

Recent research has begun exploring the integration of social media sentiment into cryptocurrency prediction models. Studies analyzing Twitter sentiment have found positive correlations between tweet volumes, sentiment scores, and Bitcoin price movements. Reddit discussions, particularly from cryptocurrency-focused subreddits, have also been identified as valuable sentiment sources. Natural Language Processing techniques including lexicon-based methods and supervised learning classifiers have been employed for sentiment extraction.

Advanced approaches utilize transformer-based models such as BERT and RoBERTa for sentiment analysis, achieving superior accuracy in capturing nuanced opinions and contextual meanings. Multimodal fusion techniques that combine sentiment features with price data have shown promising results. However, challenges remain in handling noisy social media data, identifying influential users, managing real-time data streams, and developing robust models that generalize across different cryptocurrencies and market conditions.

## III. PROPOSED METHODOLOGY

The proposed cryptocurrency price prediction system consists of five main components: data collection, sentiment analysis, feature engineering, predictive modeling, and evaluation. The system architecture enables real-time processing of social media data while maintaining high prediction accuracy across multiple cryptocurrencies.

### A. Data Collection

Our data collection pipeline aggregates information from multiple sources including Twitter, Reddit, Telegram, and specialized cryptocurrency forums. For Twitter, we collect tweets containing cryptocurrency-related hashtags and mentions using the Twitter API, filtering for influential accounts and high-engagement content. Reddit data is extracted from cryptocurrency subreddits including r/cryptocurrency, r/bitcoin, and r/ethereum, capturing posts, comments, and voting scores. Telegram messages from public cryptocurrency channels are collected through the Telegram Bot API.

Historical price data is obtained from cryptocurrency exchanges through public APIs, including OHLCV data at hourly intervals. Technical indicators such as moving averages, Relative Strength Index, MACD, and Bollinger Bands are calculated from price data. The data collection system operates continuously, storing raw data in a distributed database for subsequent processing. Data preprocessing includes duplicate removal, spam filtering, language detection, and temporal alignment between social media posts and price data.

### B. Sentiment Analysis

Sentiment analysis employs a multi-stage pipeline combining traditional NLP techniques with modern transformer-based models. Text preprocessing includes tokenization, lowercase conversion, URL removal, emoji handling, and cryptocurrency-specific abbreviation expansion. We utilize FinBERT, a BERT model fine-tuned on financial text, to classify sentiment as positive, negative, or neutral. The model achieves 89% accuracy on cryptocurrency-specific sentiment classification tasks.

For each social media post, we extract multiple sentiment features including polarity score, subjectivity score, emotional intensity, and topic relevance. Cryptocurrency mentions are identified and linked to specific coins, enabling coin-specific sentiment aggregation. Temporal sentiment features are computed by aggregating sentiment scores over various time windows including 1-hour, 6-hour, 24-hour, and 7-day periods. We also calculate sentiment momentum and volatility metrics to capture changing public opinion dynamics.

### C. Feature Engineering

Feature engineering combines sentiment-based features with traditional market indicators to create a comprehensive feature set for prediction. Market features include historical prices, trading volumes, price returns, and 20 technical indicators computed at multiple time scales. Sentiment features encompass aggregated sentiment scores, post volumes, engagement metrics, influential user activity, and sentiment divergence measures across different platforms.

Advanced features capture the relationship between sentiment and price movements, including sentiment-price correlation coefficients, sentiment change rates, and divergence between sentiment trends and price trends. User influence weights are incorporated by assigning higher importance to posts from verified accounts, high-follower users, and historically accurate predictors. The final feature set comprises 85 features normalized using standard scaling to ensure equal contribution to the prediction model.

### D. Predictive Modeling

The predictive modeling component employs two complementary approaches: LSTM neural networks and Gradient Boosting machines. The LSTM architecture consists of three LSTM layers with 128, 64, and 32 units respectively, followed by dropout layers for regularization and dense layers for final prediction. The network processes sequences of 168 hours as input, leveraging both historical price patterns and time-aligned sentiment features. We use Adam optimizer with learning rate scheduling and early stopping to prevent overfitting.

Gradient Boosting models implemented using XGBoost provide an alternative approach that excels at capturing complex nonlinear relationships between features and target prices. We configure the model with 500 trees, maximum depth of 8, and learning rate of 0.05. The model is trained using temporal cross-validation to ensure robust performance across different market conditions. Feature importance analysis reveals that recent sentiment scores, price momentum, and trading volume rank among the most predictive features.

An ensemble approach combines LSTM and Gradient Boosting predictions using weighted averaging, where weights are determined through validation set performance. The LSTM model receives 60% weight for its superior long-term trend prediction, while Gradient Boosting receives 40% weight for capturing short-term fluctuations. This ensemble framework provides robust predictions across multiple time horizons ranging from 1-hour to 7-day forecasts.

## IV. EXPERIMENTAL SETUP AND RESULTS

### A. Dataset

Our dataset spans 18 months from January 2023 to June 2024, covering three major cryptocurrencies: Bitcoin, Ethereum, and Litecoin. The social media corpus contains 5.2 million tweets, 850,000 Reddit posts and comments, and 420,000 Telegram messages. Price data includes hourly OHLCV values from major exchanges. The dataset is split temporally with 70% for training, 15% for validation, and 15% for testing, ensuring the model is evaluated on future unseen data that simulates real-world deployment.

### B. Evaluation Metrics

Model performance is assessed using multiple metrics including Mean Absolute Error, Root Mean Squared Error, Mean Absolute Percentage Error, and directional accuracy. MAE measures average prediction error in absolute terms, while RMSE penalizes larger errors more heavily. MAPE provides percentage-based error metrics facilitating comparison across different cryptocurrencies. Directional accuracy indicates the percentage of correct price movement predictions, which is crucial for trading applications.

### C. Performance Analysis

Table I presents the performance comparison between baseline models using only price data and our proposed models incorporating sentiment features. For Bitcoin 7-day prediction, the baseline LSTM achieves 4.8% MAPE, while the sentiment-enhanced LSTM reduces this to 3.2%, representing a 33% improvement. Similar improvements are observed across Ethereum and Litecoin predictions, with MAPE reductions of 18% and 22% respectively.

Cryptocurrency	Model	MAPE (%)	Direction Acc (%)
Bitcoin	Baseline LSTM	4.8	68.5
Bitcoin	<b>Sentiment-LSTM</b>	<b>3.2</b>	<b>76.8</b>
Ethereum	Baseline LSTM	5.5	65.2
Ethereum	<b>Sentiment-LSTM</b>	<b>4.5</b>	<b>72.1</b>
Litecoin	Baseline LSTM	6.2	62.8
Litecoin	<b>Sentiment-LSTM</b>	<b>4.8</b>	<b>69.5</b>

**TABLE I PREDICTION PERFORMANCE ACROSS CRYPTOCURRENCIES**

Directional accuracy improvements are equally significant, with sentiment-enhanced models achieving 76.8% accuracy for Bitcoin compared to 68.5% for baseline models. This 8.3 percentage point improvement demonstrates the value of sentiment features for predicting price movement direction, which is crucial for trading strategies. The ensemble model combining LSTM and Gradient Boosting achieves the best overall performance with 3.0% MAPE and 78.2% directional accuracy for Bitcoin predictions.

### D. Sentiment Impact Analysis

Feature importance analysis reveals that sentiment features account for approximately 35% of the model's predictive power. Twitter sentiment scores rank as the third most important feature after recent price changes and trading volume. Reddit sentiment demonstrates particularly strong predictive value during market corrections and rallies. Temporal sentiment dynamics, especially rapid sentiment shifts, serve as leading indicators for significant price movements occurring 6-12 hours later.

Platform-specific analysis shows that Twitter sentiment correlates most strongly with short-term price movements within 24 hours, while Reddit discussions provide better signals for medium-term trends spanning 3-7 days. Telegram sentiment exhibits higher noise levels but captures early signals from trading communities. The combination of multi-platform sentiment provides complementary information that enhances prediction robustness across different time horizons.

## V. DISCUSSION

The experimental results validate the hypothesis that social media sentiment significantly enhances cryptocurrency price prediction. The 15-22% improvement in MAPE across different cryptocurrencies demonstrates consistent value across market conditions and asset characteristics. The particularly strong performance during volatile periods suggests that sentiment analysis effectively captures market psychology and collective behavior that drives rapid price movements.

Several factors contribute to the effectiveness of sentiment-based prediction. First, cryptocurrency markets remain largely sentiment-driven compared to traditional financial markets, with limited fundamental valuation metrics. Second, social media platforms serve as primary information channels for cryptocurrency communities, creating tight coupling between online discussions and market activity. Third, the decentralized and global nature of cryptocurrency trading means that sentiment trends can quickly translate into coordinated buying or selling pressure.

Challenges identified during this research include handling spam and bot-generated content that can distort sentiment signals, identifying and weighing influential users appropriately, managing the computational overhead of real-time sentiment analysis at scale, and maintaining model performance during unprecedented market events not represented in training data. Bot detection algorithms and user credibility scoring mechanisms partially address these challenges but require ongoing refinement.

The system demonstrates practical viability for integration into trading platforms and portfolio management tools. Real-time sentiment monitoring combined with predictive models can support automated trading strategies, risk management systems, and investment decision support. Privacy and ethical considerations must be addressed, particularly regarding the use of user-generated content and potential market manipulation through coordinated sentiment campaigns.

## VI. CONCLUSION

This paper presents a comprehensive cryptocurrency price prediction system that integrates social media sentiment analysis with traditional market indicators. The proposed approach combines advanced Natural Language Processing techniques for sentiment extraction with LSTM networks and Gradient Boosting models for price forecasting. Experimental evaluation demonstrates 15-22% improvement in prediction accuracy compared to baseline models, with the ensemble approach achieving 3.2% MAPE for Bitcoin 7-day predictions.

The research confirms that social media sentiment serves as a valuable leading indicator for cryptocurrency price movements, particularly during periods of high volatility. Multi-platform sentiment aggregation from Twitter, Reddit, and Telegram provides complementary signals that enhance prediction robustness. Feature importance analysis reveals that sentiment features contribute approximately 35% of predictive power, ranking among the top features alongside price momentum and trading volume.

Future research directions include investigating attention mechanisms for weighting influential users dynamically, developing adversarial training methods to improve robustness against market manipulation, exploring multimodal analysis incorporating images and videos from social media, and extending the framework to predict additional market metrics such as volatility and liquidity. Integration with blockchain analytics and on-chain metrics could provide additional predictive signals. Real-time deployment and continuous learning mechanisms will enable the system to adapt to evolving market dynamics and emerging cryptocurrencies.

## VI. ACKNOWLEDGMENT

The authors would like to thank the Engineering College for providing computational infrastructure and research support. We acknowledge cryptocurrency exchanges and social media platforms for providing access to data through their public APIs. Special thanks to the Department of Data Science faculty members and industry advisors for their valuable insights and guidance throughout this research project.

## REFERENCES

- [1] S. McNally, J. Roche and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning", Proc. Euromicro Int. Conf. Parallel, Distributed and Network-based Processing, pp. 339-343, 2018.
- [2] 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, pp. 112395, 2020.
- [3] A. Patel, R. Pasad and S. Vaghela, "Cryptocurrency Price Prediction Using Neural Network and Recurrent Neural Network", Proc. Int. Conf. Inventive Communication and Computational Technologies, pp. 1394-1399, 2019.
- [4] M. A. Kraaijeveld and J. De Smedt, "The Predictive Power of Public Twitter Sentiment for Forecasting Cryptocurrency Prices", Journal of International Financial Markets, Institutions and Money, vol. 65, pp. 101188, 2020.
- [5] T. Stenqvist and J. Lonnö, "Predicting Bitcoin Price Fluctuation with Twitter Sentiment Analysis", KTH Royal Institute of Technology, Master's Thesis, 2017.
- [6] J. Abraham, D. Higdon, J. Nelson and J. Ibarra, "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis", SMU Data Science Review, vol. 1, no. 3, pp. 1-22, 2018.
- [7] E. Alessandretti, A. ElBahrawy, L. M. Aiello and A. Baronchelli, "Anticipating Cryptocurrency Prices Using

Machine Learning", Complexity, vol. 2018, pp. 1-16, 2018.

[8] S. Lahmiri and S. Bekiros, "Cryptocurrency Forecasting with Deep Learning Chaotic Neural Networks", Chaos, Solitons & Fractals, vol. 118, pp. 35-40, 2019.

[9] D. Shah and K. Zhang, "Bayesian Regression and Bitcoin", Proc. Allerton Conf. Communication, Control, and Computing, pp. 409-414, 2014.

[10] A. Burnie and E. Yilmaz, "Social Media and Bitcoin Metrics: Which Words Matter", Royal Statistical Society Journal, vol. 182, no. 2, pp. 567-594, 2019.

[11] Y. B. Kim, J. G. Kim, W. Kim, J. H. Im, T. H. Kim, S. J. Kang and C. H. Kim, "Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies", PLoS ONE, vol. 11, no. 8, pp. e0161197, 2016.

[12] I. Madan, S. Saluja and A. Zhao, "Automated Bitcoin Trading via Machine Learning Algorithms", Stanford University Technical Report, pp. 1-5, 2015.

[13] Napoleon, D., et al. "An efficient numerical method for the prediction of clusters using k-means clustering algorithm with bisection method." International Conference on Computing and Communication Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details