

## Review of Sentiment Analysis in Cryptocurrency Trading

Dr. Keerthi Kumar H M<sup>1</sup>, M Vishwas Bharadwaj<sup>2</sup>, P Shreyas Gowda<sup>3</sup>, Sanjana Desh<sup>4</sup>, U N Poorvi Vasishta<sup>5</sup>

<sup>1</sup>Associate professor, Department of CSE, Malnad College of Engineering, Hassan, Karnataka, India.

<sup>2,3,4,5</sup>UG Scholar, Dept. of CSE, Malnad College of Engg, Hassan, Karnataka, India.

**Emails:** [hmk@mcehassan.ac.in](mailto:hmk@mcehassan.ac.in)<sup>1</sup>, [vishwasmb.hassan@gmail.com](mailto:vishwasmb.hassan@gmail.com)<sup>2</sup>, [shreyasgowdap2004@gmail.com](mailto:shreyasgowdap2004@gmail.com)<sup>3</sup>, [aleena.selan@gmail.com](mailto:aleena.selan@gmail.com)<sup>4</sup>, [pvun26@gmail.com](mailto:pvun26@gmail.com)<sup>5</sup>

### Abstract

*The rapid rise of cryptocurrencies has impacted the global socio-economic landscape, encouraging investors to seek income through crypto trading. Due to the market's volatility and complex interdependencies, researchers have built various prediction models using machine learning, deep-learning, and sentiment-based hybrid algorithms. Notably, the DLCFS (Deep Learning Cryptocurrency Forecasting considering Sentiment) framework incorporates market features, trading volume, and sentiment from Reddit to improve price predictions for Bitcoin, Ethereum, and Litecoin—achieving high accuracy when compared to traditional machine learning models. Alongside forecasting, sentiment analysis plays an important role in understanding market trends and investor behavior. With growing user-generated content across different platforms like social media and news sites, extracting public sentiment through NLP has become essential. Recent works explore advanced models and datasets tailored to the unique linguistic features of crypto-related content, highlighting the need for robust and adaptive sentiment analysis techniques in this dynamic domain.*

**Keywords:** Cryptocurrency Market Prediction, Sentiment Analysis, Natural Language Processing (NLP), Deep Learning Models, Support Vector Machine (SVM), Social Media Mining, LSTM Networks, Reddit and Twitter Sentiment, Hybrid Forecasting Models, DLCFS Framework.

### 1. Introduction

Cryptocurrency is a digital currency [8] that was designed to replace traditional currency and has been influencing the perceptions of communities in the generation nowadays. Cryptocurrencies differ from modern currencies as modern currencies are regulated in the sense of whether financial transactions are executed through the involvement of a third-party organization such as a bank. Bitcoin, being the first cryptocurrency introduced in 2009, used a Proof of Work algorithm (PoW) to ensure system integrity and consistency. Nowadays, newer digital currencies, namely Ethereum, BNB, Cardano and many more, have adopted a variety of algorithms such as Proof-of-Stake (PoS) to reduce carbon footprints. The cryptocurrency was designed to replace the current centralized financial system with a more transparent, secure, and distributed decentralized system. In the current market, 8,685 cryptocurrencies are being developed and traded

actively and the global market cap for cryptocurrency is at 1.023 trillion USD (until February 2023). Cryptocurrency, which emerged in 2008 as a decentralized digital currency [1], has attracted worldwide people attention due to its rapid expansion and investment opportunities. However, the volatile and high-risk nature of cryptocurrency markets complicates decision-making for traders. Social media platforms, in particularly Twitter, serve as a significant source of public sentiment, where opinions conveyed through tweets and retweets can impact trading actions. Sentiment analysis is used to classify these views as positive, negative, or neutral categories, providing useful information about market sentiment. Among various methodologies, machine learning—especially Support Vector Machines (SVM) [2]—has demonstrated considerable accuracy in sentiment classification. Furthermore, feature selection methods such as Chi-

square improve performance by eliminating irrelevant data. This study investigates sentiment analysis of cryptocurrency-related tweets and retweets utilizing SVM and Chi-square, with the objective of aiding traders in making more informed investment choices. In the contemporary digital landscape, vast amounts of text are generated via social media, emails, reviews, and news articles. It is essential to extract insights from this unstructured data across various sectors. Text analysis, also known as Natural Language Processing (NLP), converts raw text into structured data to identify patterns and trends. Machine learning integration enhances this process by facilitating the automated comprehension and categorization of language. A significant application of this technology is sentiment analysis, which explains the emotional tone of text if it is positive, neutral, or negative providing valuable insights for fields such as marketing, finance, and public opinion research.

## 2. Literature Survey

Cryptocurrency is a digital, decentralized currency designed to function without the control of centralized financial institutions. In contrast to traditional currencies, which depend on intermediaries like banks to verify transactions, cryptocurrencies utilize blockchain technology and consensus algorithms such as Proof of Work (PoW) and Proof of Stake (PoS) to maintain integrity and transparency. With thousands of cryptocurrencies currently in circulation and a market cap reaching over a trillion USD, the dynamic and volatile nature of this financial ecosystem has driven interest in predictive and analytical tools, particularly those powered by artificial intelligence and machine learning. One significant area of research involves the use of sentiment analysis to know how public opinion especially from social-media platform like Twitter—affects cryptocurrency market trends. Hasan, Oetama, and Saonard [2], [4] focused on this by employing a machine learning framework involving Support Vector Machines (SVM) along with Chi-square feature selection. Tweets and retweets collected from October to December 2021 were manually labeled by language experts. The data underwent comprehensive preprocessing, including tokenization, stemming, lemmatization, and

duplicate removal, followed by polarity and subjectivity scoring using the TextBlob library. Text was vectorized using TF-IDF, and the Chi-square technique was applied to select the most relevant features, ultimately improving classification accuracy. Their structured pipeline demonstrates a robust approach to extracting meaningful sentiment from noisy, unstructured data. Beyond conventional machine learning, models on deep learning have also shown promise in sentiment classification tasks. Michael Nair et al. [5] provided a comparative analysis between supervised learning methods and deep learning techniques, showing the enhanced ability of the latter to capture semantic and contextual nuances in text. The main methodologies mentioned in this paper are Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU) and CONV1D (1-Dimensional Convolutional Neural Network). Similarly, Roumeliotis et al. [6] compared deep learning architectures such as RNNs and LSTMs against classical models such as Support Vector Machines (SVM) and Random Forest (RF), concluding that deep models perform significantly better in capturing time-sequential sentiment patterns. Jahanbin and Chahooki [7] emphasized the importance of transfer learning and the application of pre-trained language models such as Bidirectional Encoder Representations from Transformers (BERT), which further improve sentiment classification by understanding context and sarcasm in text better than traditional methods. Integrating sentiment with historical pricing data has also evolved as a key strategy. Jia Ming Low et al. [8] introduced deep learning framework called DLCFS (Deep Learning Combined Feature Set) that utilizes LSTM networks to predict cryptocurrency prices using both historical price data and Reddit-derived sentiment scores. Their findings showed that the combination of sentiment and financial indicators significantly improves prediction accuracy, outperforming standalone time-series models like ARIMA and traditional LSTM models. This underscores the value of incorporating community-driven sentiment in financial forecasting. Building upon this line of investigation, various research efforts have looked into the combination of social

media sentiment and cryptocurrency market forecasting using both conventional and deep learning techniques. Raheman et al. [11] evaluated more than twenty sentiment analysis models on a dataset comprising Twitter and Reddit posts related to cryptocurrencies, ultimately determining that the Aigents model was the most effective due to its interpretable design and customizable n-gram vocabulary. Adjusting the model for specific sentiment within the domain led to a notable enhancement in predictive correlation with fluctuations in Bitcoin prices, increasing from 0.33 to 0.57. Huang et al. [13], concentrating on the Chinese market, introduced a deep learning strategy that utilized LSTM networks trained on Sina-Weibo data, augmented with a manually developed sentiment dictionary specific to cryptocurrency. Their model surpassed a standard autoregressive time-series method by 18.5% in precision and 15.4% in recall, demonstrating LSTM's adaptability to complex, non-English sentiment data. In contrast, previous research by Neri et al. [12] addressed broader uses of sentiment analysis in media monitoring. Although not specifically aimed at financial forecasting, their study utilized a semantic-driven engine on Facebook postings to relate audience sentiment to the performance of traditional news media, providing important methodological contributions to multilingual sentiment extraction and real-time opinion monitoring. Together, these studies emphasize the significance of both linguistic

customization and model interpretability in boosting the predictive strength of sentiment-based financial analytics. In summary, the literature reveals an undeniable move away from more conventional statistical models towards hybrid machine learning and deep learning approaches. These models not only study the price trends but also incorporate public sentiment, enabling a more holistic understanding of cryptocurrency market dynamics. The integration of social media sentiment—particularly from platforms like Twitter and Reddit—with financial data is proving essential in enhancing prediction models, offering traders and analysts better tools for decision-making in an inherently volatile market.

### 3. Dataset

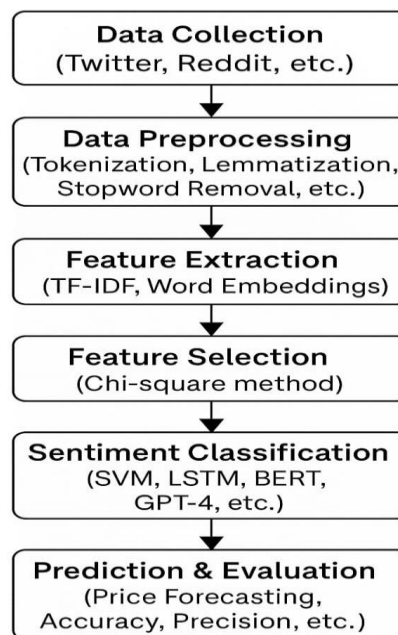
A thorough summary of the datasets used in the research covered in this paper is provided in the table below. The various sentiment analysis approaches used on cryptocurrency markets are based on these datasets. They present a wide variety of user-generated content and sentiment-rich data, collected from sites like Twitter, Reddit, Facebook, Sina-Weibo, and financial APIs. The size, distribution of sentiment classes, language, and accessibility of each dataset vary. This overview highlights the data foundations supporting several machine learning and deep learning methods applied in cryptocurrency sentiment analysis by providing a summary of the datasets cited in previous research.

**Table 1 Dataset**

Dataset Name	Instances	Class	Availability
Twitter Api[2]	NA	Positive Negative	Twitter crawler via the Twitter API
Kaggle[4]	NA	Positive Neutral Negative	
Twitter BTC Tweets Dataset[5]	1.5 million tweets	<b>Positive:</b> 2,00,000 <b>Negative:</b> 2007 <b>Neutral:</b> 26,00	Kaggle ( <a href="https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets">https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets</a> )
Tweet Data from Twitter API (2022)[5]	154,481 articles, 570,865 tweets, 90,268 Telegram posts	Positive Neutral Negative	Twitter API and open sources

Crypto News + dataset compiled by Oliviervha[6]	Approximately 31,037 rows	<b>Positive:</b> 1,666 <b>Neutral:</b> 1,666 <b>Negative:</b> 1,666	Publicly accessible on Kaggle under the Database Contents License (DbCL) v1.0 <a href="https://www.kaggle.com/dataset/s/oliviervha/crypto-news">https://www.kaggle.com/dataset/s/oliviervha/crypto-news</a>
SemEval-2015 Task 12 & SemEval-2016 Task 5[7]	29,860 cleaned tweets	<b>Positive:</b> 7,215 <b>Neutral:</b> 3,995 <b>Negative:</b> 5,302	Publicly available via SemEval
Yahoo Finance API[7]	NA	Various crypto currencies	Collected directly via APIs and online sources, not publicly available
Twitter and Reddit Feeds (2021)[9]	~100,000 posts (tweets + Reddit posts) over 6 months (Jul–Dec 2021); 490 manually labeled tweets/posts for evaluation	Positive Negative Contradictive	Public posts via official Twitter and Reddit APIs
Facebook Posts (Media Sentiment)[10]	~1000 Facebook posts comparing Rai and La7 news channels	Positive Negative	Crawled from Facebook; not explicitly stated as public
Sina-Weibo Posts (Chinese social media)[11]	24,000 original posts + 70,000 comments over 8 days	Positive Neutral Negative (manually labeled)	Crawled from Sina-Weibo using web scraper

#### 4. Methodology



**Figure 1 Methodology**

## 5. Comparative Analysis

**Table 2 Comparative Analysis**

Author	Methodology	Dataset	Precision	Recall	F1-Score	Accuracy
I. D.Hasan, R. S. Oetama and A. L. Saonard[2]	SVM	Twitter API	68%	69%	63%	69%
S. Bhatt, M. Ghazanfar, and M. H. Amirhosseini[4]	SVC Logistic Regression Naïve bayes KNN XGBoost Multi Modal Fusion	Kaggle	75.7%		70.1%	75.7% 76.2% 75.8% 59.6% 61.3% 85%
M. Nair,L.A. Abd-Elmegid, and M.I. Marie[5]	LSTM	Bitcoin tweets from Kaggle	88%	82%	95.95%	95.95%
M. Nair,L.A. Abd-Elmegid, and M.I. Marie[5]	RNN	Bitcoin tweets from Kaggle	29%	33%	80.95%	80.59%
M. Nair,L.A. Abd-Elmegid, and M.I. Marie[5]	GRU	Bitcoin tweets from Kaggle	90%	80%	95.82%	95.82%
M. Nair,L.A. Abd-Elmegid, and M.I. Marie[5]	Bi-LSTM+ CONV1D	Bitcoin tweets from Kaggle	90%	79%	95.67%	95.67%
M. Chakraborty and S. Subramaniam[6]	Base GPT-4	Crypto News + dataset	Negative: 85.5% Positive: 80.3% Neutral : 80.1%	Negative: 85.0% Positive: 79.6% Neutral: 84.1%	Negative: 84.1% Positive: 82.4% Neutral: 82.2%	82.9%
M. Chakraborty and S. Subramaniam[6]	Fine-tuned GPT-4	Crypto News + dataset	Negative : 87.3% Positive: 88.8% Neutral : 84.2%	Negative : 86.5% Positive: 85.9% Neutral: 87.7%	Negative : 85.9% Positive: 87.3% Neutral: 85.9%	86.7%

M. Chakraborty and S. Subramaniam[6]	BERT(Fine-tuned )	Crypto News + dataset	Negative : 79.7% Positive: 84.9% Neutral : 85.6%	Negative : 79.3% Positive: 82.6% Neutral: 83.5%	Negative : 81.3% Positive: 84.5% Neutral: 83.7%	83.3%
M. Chakraborty and S. Subramaniam[6]	FineBERT (Fine-tuned)	Crypto News + dataset	Negative : 86.3% Positive: 80.4% Neutral : 86.5%	Negative : 82.9% Positive: 85.0% Neutral: 85.0%	Negative : 82.6% Positive: 85.8% Neutral: 82.6%	84.3%
Jahanbin K. and Chahooki M. A. Z.[7]	HDRB (Hybrid Deep neural network of RoBERTa and Bidirectional Gated Recurrent Unit)	SemEval-2015 Task 12	88.63%	87.30%	87.96%	77.55%
		SemEval-2016 Task 5	83.53%	82.21%	82.86%	77.35%
Jahanbin K. and Chahooki M. A. Z.[7]	BGA model	Yahoo Finance API	NA	NA	Bitcoin: 86.14% Ethereum: 82.81% Binance: 81.24% Ripple: 80.06% Dogecoin: 74.82% Solana: 70.23% Cardona: 74.82%	Bitcoin: 87.25 Ethereum: 86.41 Binance: 82.46 Ripple: 80.95 Dogecoin: 78.35% Solana: 73.75% Cardona: 78.35%
Jia Ming Low, Zi Jian Tan, Tiong Yew Tang and Narishah Mohamed Salleh[8]	DLCFS	Historical Price Data + Reddit Submissions (Bitcoin, Ethereum, Litecoin)				Bitcoin: 99.18%, Ethereum: 99.05%, Litecoin: 96.82%

A.Raheman, A. Kolonin, I. Fridkins, I. Ansari, and M. Vishwas[9]	Aigents (n-gram based, interpretable model) + Fine tuned	Twitter & Reddit (100k posts); 490 labeled for eval				Pearson Correlation: 0.57 (Aigents+); 0.33 (original)
F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By [10]	Supervised & Unsupervised Classification; Semantic NLP System	~1000 Facebook posts (Rai vs. La7)	>87%	>93%		Not explicitly reported
X. Huang, W. Zhang, X. Tang, M. Zhang, J. Surbiryala, V. Iosifidis, Z. Liu, and J. Zhang[11]	LSTM (deep learning, custom crypto sentiment dictionary)	24,000 Weibo posts + 70,000 comments	87%	92.5		Outperformed AR model by 18.5% precision and 15.4% recall

## Conclusion

The survey of recent studies reveals a clear progression in the techniques used for sentiment analysis in cryptocurrency trading, moving from conventional machine learning models to sophisticated deep learning architectures. Support Vector Machines (SVM) paired with Chi-square feature selection showed reliable performance on Twitter-based sentiment datasets, achieving up to 69% accuracy. However, the deep learning methods have consistently demonstrated superior effectiveness. Notably, the LSTM-based model by Huang et al. achieved 87% precision and 92.5% recall on Chinese-language Weibo posts, showcasing the advantage of recurrent networks in modeling sequential sentiment data. Likewise, the DLCFS framework integrated Reddit sentiment with historical price indicators, yielding accuracy rates as high as 99.18% for Bitcoin. Raheman et al.'s work using the Aigents model highlighted the importance of model interpretability and domain-specific tuning, reaching a Pearson correlation of 0.57 with market trends. Despite these achievements, multiple areas remain open for further research and development.

Future efforts should focus on multilingual sentiment modeling and cross-platform sentiment integration, given the global nature of cryptocurrency discussions. Additionally, real-time data fusion of sentiment signals with high-frequency market data could significantly improve responsiveness and prediction accuracy. A particularly important challenge is the detection of sarcasm and figurative language, which often misleads traditional sentiment models. Incorporating sarcasm detection mechanisms, possibly through advanced transformers or fine-tuned language models, could substantially enhance sentiment reliability. Moreover, further exploration of transfer learning and large language models (LLMs) such as BERT, GPT, or their fine-tuned variants may provide better context-awareness and adaptability across domains. As a follow-up to the sentiment analysis performed in this study, future endeavours could look into the predictive capabilities of price movements. The work done by Haritha and Sahana [12] showcased the effectiveness of BERT-based sentiment analysis combined with GRU for Bitcoin price predictions,

achieving remarkable accuracy. Mondal et al. [13] argued for employing sentiment-informed narratives associated with certain structural price movements. Under certain parameters, Kang et al. [14] analysed the predictive ability of Korean news sentiment concerning short-term returns and found it relevant. Alghamdi et al. [15] demonstrated that sentiment analysis using SVM, combined with LSTM-based price forecasting, can reveal a correlation between user sentiment and price volatility in cryptocurrencies such as BTC and ETH. These studies suggest that with appropriate sentiment modelling, it can be seamlessly integrated as a fundamental component in market prediction frameworks. In conclusion, hybrid approaches that integrate sentiment analysis with deep learning and financial indicators have emerged as the most promising tools for cryptocurrency trend prediction. As the field evolves, developing context-aware, real-time, and ethically robust sentiment analysis systems—especially those capable of handling sarcasm and linguistic nuances—will be critical for supporting informed decision-making in this highly volatile market.

## References

- [1]. S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: <https://bitcoin.org/bitcoin.pdf>
- [2]. D. Hasan, R. S. Oetama and A. L. Saonard, "Sentiment Analysis on Cryptocurrency Based on Tweets and Retweets Using Support Vector Machines and Chi-Square," 2022 7th International Conference on Informatics and Computing (ICIC), Tangerang, Indonesia, 2022, pp. 1–6, doi: 10.1109/ICIC56845.2022.10006895
- [3]. K. Karthika, "An efficient approach for sarcasm detection in tweets using polarity flip," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 5, no. 2, pp. 1822–1825, Apr. 2019.
- [4]. S. Bhatt, M. Ghazanfar, and M. H. Amirhosseini, "Sentiment-Driven Cryptocurrency Price Prediction: A Machine Learning Approach Utilizing Historical Data and Social Media Sentiment Analysis," University of East London, 2023.
- [5]. Michael Nair\*, Laila A. Abd-Elmegid, and Mohamed I. Marie, "Sentiment analysis model for cryptocurrency tweets using different deep learning techniques", *Journal of Intelligent Systems*, Dec. 2024.
- [6]. Roumeliotis, K.I.; Tselikas, N.D.; Nasiopoulos, D.K, "LLMs and NLP Models in Cryptocurrency Sentiment Analysis: A Comparative Classification Study", *Big Data Cogn. Comput.* 2024, 8, 63. <https://doi.org/10.3390/bdcc8060063>.
- [7]. K. Jahanbin, M. A. Z. Chahooki, "Aspect-Based Sentiment Analysis of Twitter Influencers to Predict the Trend of Cryptocurrencies Based on Hybrid Deep Transfer Learning Models", *IEEE Access*, Nov. 2023.
- [8]. Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction Jia Ming Low, Zi Jian Tan, Tiong Yew Tang and Narishah Mohamed Salleh Sunway Business School, Sunway University, 47500 Subang Jaya, Selangor, Malaysia. 2023
- [9]. Raheman, A. Kolonin, I. Fridkins, I. Ansari, and M. Vishwas, "Social Media Sentiment Analysis for Cryptocurrency Market Prediction," arXiv preprint, arXiv:2204.10185v1, Apr. 2022. [Online]. Available: <https://arxiv.org/abs/2204.10185>
- [10]. F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, "Sentiment Analysis on Social Media," in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012, pp. 919–926. doi: 10.1109/ASONAM.2012.164.
- [11]. X. Huang, W. Zhang, X. Tang, M. Zhang, J. Surbiryala, V. Iosifidis, Z. Liu, and J. Zhang, "LSTM Based Sentiment Analysis for Cryptocurrency Prediction," arXiv preprint, arXiv:2103.14804v4, Oct. 2021. [Online]. Available: <https://arxiv.org/abs/2103.14804>
- [12]. Haritha G. B. and Sahana N. B., "Cryptocurrency Price Prediction Using Twitter Sentiment Analysis," in *Proc. CS & IT - CSCP 2023*, pp. 13–22, 2023. doi: 10.5121/csit.2023.130302
- [13]. Lubdhak Mondal, Udeshya Raj, Abinandhan



- S, Began Gowsik S, Sarwesh P, and Abhijeet Chandra, "Causality between Sentiment and Cryptocurrency Prices," arXiv preprint arXiv:2306.05803, Jun. 2023. [Online]. Available: <https://arxiv.org/abs/2306.05803>
- [14]. Ryeomyung Kang, Soosung Hwang, Jinho Shin, and Yeeun Yoo, "Cross-border Interactions of Cryptocurrency Prices and News Sentiment: Who Drives Whom?," SSRN Electronic Journal, SSRN:4714113, 2024. [Online]. Available: <https://ssrn.com/abstract=4714113>
- [15]. S. Alghamdi, S. Alqethami, T. Alsubait, and H. Alhakami, "Cryptocurrency Price Prediction Using Forecasting and Sentiment Analysis," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 13, no. 10, pp. 891–900, 2022. [Online]. Available: [https://thesai.org/Downloads/Volume13No10/Paper\\_105Cryptocurrency\\_Price\\_Prediction\\_using\\_Forecasting\\_and\\_Sentiment\\_Analysis.pdf](https://thesai.org/Downloads/Volume13No10/Paper_105Cryptocurrency_Price_Prediction_using_Forecasting_and_Sentiment_Analysis.pdf)