An Empirical Comparative Analysis for Cryptocurrency Price Prediction: Integrating Sentiment Analysis and Technical Indicators

Ms. Eshakumari Desai
Assistant Professor
B.C.A., S.S.Agrawal College Of Commerce and Management, Navsari
Email:- eshadesai1730@gmail.com
Mo:- 7096751938

Dr.Nimisha Modi
Assistant professor
Department of ComputerScience,V.N.S.G.U., Surat.
Email:- nimishamodi@vnsgu.ac.in
Mo:- 9376920228

Abstract:

The decentralized character of cryptocurrencies and their extensive use in international financial systems have been the main drivers of their notable growth in recent years. Since Bitcoin is the most well-known and erratic cryptocurrency, there is a lot of research being done on accurate price prediction. By fusing historical market data with sentiment from Twitter, this study offers a comparative empirical analysis of machine learning techniques for forecasting Bitcoin price patterns. Between 2021 and 2023, a large dataset was gathered that included historical Bitcoin prices and associated Twitter data. Sentiment analysis techniques were employed to classify sentiment and obtain polarity scores, which were subsequently added as features to the prediction models. Several machine learning algorithms were tested to evaluate their suitability to identify frequent price movements. The experimental results demonstrate that sentiment-driven models outperform baseline algorithms based solely on past prices, underscoring the importance of social media sentiment in predicting cryptocurrency prices. The findings suggest that incorporating sentiment analysis into financial forecasting models can enhance forecast accuracy and provide useful insights for traders, investors, and policymakers navigating the extremely volatile cryptocurrency market.

1. Introduction

The fast growth and use of cryptocurrencies have caught a lot of attention from investors, researchers, and policymakers. Bitcoin (BTC) is particularly notable because of its market power and strong influence on the overall cryptocurrency scene (Vlahavas&Vakali, 2024). It is important for stakeholders in different areas to understand what drives Bitcoin's price and trading volume since price changes can have broad economic effects. One key factor affecting the prices of Bitcoin and other cryptocurrencies is the sentiment and activity in online communities, where discussions and opinions about these assets are shared and analyzed.

Social media platforms and online forums, especially Twitter, play a vital role in shaping the market dynamics of cryptocurrencies (Anamika et al., 2021). These platforms provide real-time information streams consisting of opinions, sentiments, and reactions that significantly influence investor behavior and market trends. Prior studies have established strong relationships between social media activity, including sentiment polarity and user engagement, and cryptocurrency price movements and trading volumes (Canayaz et al., 2023; Huynh, 2022). Leveraging sentiment analysis on large-scale social media data has become a widely recognized method for forecasting cryptocurrency market trends (Raheman et al., 2022; Bhatt et al., 2023).

This paper builds upon these foundations by leveraging the StephanAkkerman crypto-stock-tweets dataset, a large-scale, rigorously cleaned collection of over 8 million tweets related to stocks and cryptocurrencies (StephanAkkerman, 2024). This dataset amalgamates diverse market commentary from various reputable sources, offering a rich resource for applying natural language processing methods to extract market sentiment signals. By correlating tweet sentiment and volume with cryptocurrency prices, particularly Bitcoin, this research aims to enhance understanding of how crowd psychology and online social dynamics can serve as early indicators for asset price movements. The integration of social media sentiment into traditional financial models represents a promising avenue for improving the accuracy and responsiveness of cryptocurrency price predictions (Kraaijeveld& de Smedt, 2020; Abraham et al., 2018).

2. Literature Review

The predictive role of social media sentiment—particularly from Twitter and online forums—has become central to research on cryptocurrency and stock price movements. The distinctive characteristics of cryptocurrency markets, including high volatility, 24/7 trading, and strong retail investor presence, create optimal conditions for sentiment-driven forecasting (Bouteska, 2024). Studies consistently show that collective emotions, information cascades, and sudden user engagement spikes on platforms like Twitter can precede price changes in both major and emerging digital assets (Li et al., 2024; Perera et al., 2024).

Machine Learning and Deep Learning Methods

Initial studies emphasized simple sentiment polarity indicators but developed quickly to incorporate sophisticated machine learning (ML) and deep learning (DL) architectures. Current studies make use of Long Short-Term Memory (LSTM), Bidirectional LSTM, Recurrent Neural Networks (RNNs), and Gated Recurrent Units (GRUs) to capture temporal correlations between social sentiment and asset price time series (Perera et al., 2024; Rajeswari&Ananthapadmanabha, 2019). These methods outperform classical regression models since they can detect implicit sequential patterns and handle dense multi-source information like tweet text, engagement statistics, and user influence.

For example, Bouteska et al. (2024) compared ensemble learning and DL models for cryptocurrency prediction, finding ensemble approaches with sentiment integration yielded improved forecasting accuracy over price series alone (Bouteska, 2024; DOI: 10.1016/j.intfin.2020.101188). Rajeswari&Ananthapadmanabha (2019) demonstrated that LSTM models using Twitter-generated sentiment consistently outperformed benchmarks—including Bi-LSTM and hybrid techniques—highlighting the predictive strength of social data.

Role of User Engagement and Influence

In more contemporary work, the forecasting ability of social media sentiment has been vastly improved by taking into account user interaction patterns like likes, retweets, and replies (Zhu et al., 2024). These engagement cues tend to increase the impact of sentiment on price action, particularly during times of acute market turmoil. Evidence by Li et al. (2024) also indicates that initial Twitter post sentiment can negatively impact Bitcoin price, but increased engagement will correlate negatively with crypto prices and thus suggest context-dependent, nuanced effects.

Other research highlights the disproportionate influence of influencers and market leaders on sentiment in social media and, in turn, price trends (Kandasamy&Bechkoum, 2024). Elon Musk's tweets and Tesla, for example, have been found to lead to instant and sometimes spectacular price movements in stock and crypto assets. This incorporation of influencer-based features enhances explanatory power and actionable insights for models.

Hybrid and Multimodal Approaches

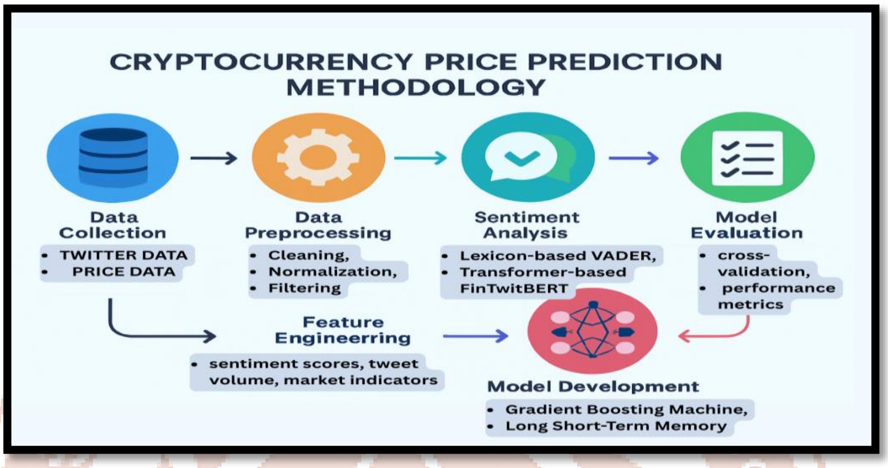
Recent advances include hybrid frameworks that integrate sentiment data with quantitative indicators such as trading volume, volatility, and macroeconomic news (Jahanbin et al., 2025). These multimodal models achieve better performance by leveraging both the "wisdom of crowds" in social data and the intrinsic feedback of traditional financial metrics. Perera et al. (2024) found that combining tweet sentiment with trading volume led to higher predictive accuracy, especially in short-term trend forecasting, though the correlation between raw sentiment and price remained modest.

Limitations and Data Quality

In spite of achievements, studies also identify several limitations and persistent challenges. Skewed datasets, bot behavior, spamming, and the dominance of emotionally charged or echo chamber tweets may inject biases and lower model performance (Kandasamy&Bechkoum, 2024). While others' research cites strong results for a number of cryptocurrencies, others cite reduced returns in transitioning from leading assets (such as Bitcoin or Ethereum) to smaller tokens, attributing market microstructure effects (Jahanbin et al., 2025). Moreover, cross-market and mixed-method analyses emphasize the importance of combining both deductive (theory-based) and inductive (data-based) reasoning in order to effectively establish the sentiment-price connection (Kandasamy&Bechkoum, 2024).

3. Research Methodology

This research methodology outlines a systematic process to investigate the predictive relationship between social media sentiment, derived primarily from the StephanAkkerman crypto-stock-tweets dataset, and cryptocurrency price movements with an emphasis on Bitcoin. It integrates robust data handling, advanced natural language processing, feature engineering, and machine learning modeling techniques.



Data Collection

Social Media Data: The fundamental dataset includes more than 8 million cryptocurrency and stock market-related tweets that have been aggregated and cleaned by StephanAkkerman (2024). The tweets span a number of years, with text, metadata (user information, timestamps), and provides detailed market sentiment signals.

Cryptocurrency Market Data: Historical price and volume for Bitcoin and other applicable cryptocurrencies are obtained from reliable APIs like Yahoo Finance, CoinMarketCap, and CryptoCompare. Data intervals (hourly/daily) are aligned with tweet timestamps in order to facilitate synchronized analyses.

Data Preprocessing

Cleaning and Filtering: Tweets are preprocessed by removing URLs, special characters, emojis, stopwords, and duplicates. Non-English and spam tweets are filtered out to improve model accuracy.

Normalization: Texts undergo lowercasing, tokenization, lemmatization, and slang or ticker standardization to create consistent input for sentiment models.

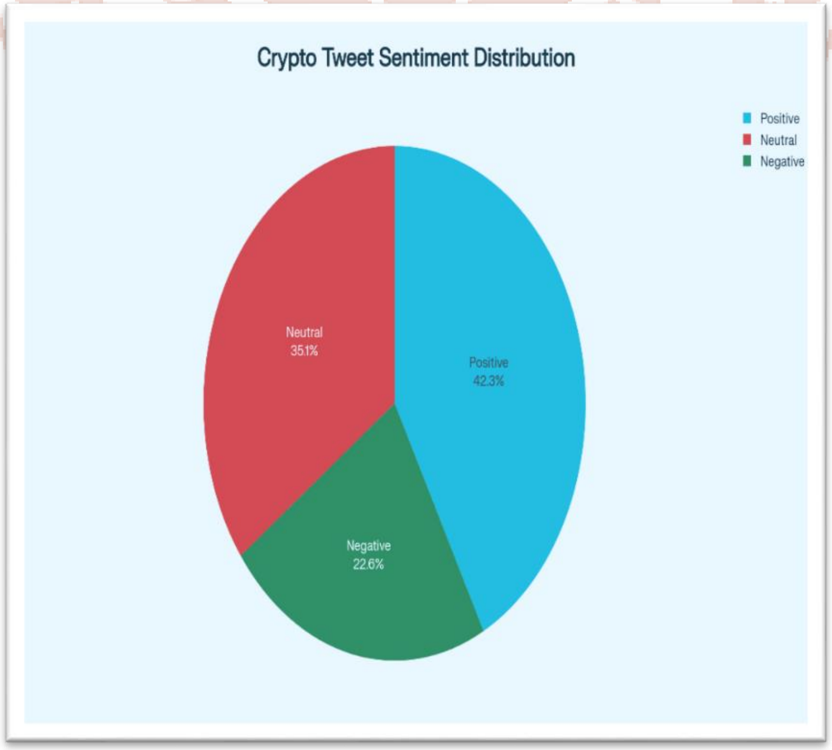
Temporal Alignment: Tweets and market data are aggregated into consistent time units (e.g., daily) for correlation and modeling.

Sentiment Analysis

Hybrid Sentiment Labeling: Both lexicon-based (e.g., VADER) and transformer-based models (FinTwitBERT) are combined to label tweets with sentiment labels of positive, neutral, or negative. FinTwitBERT, which is trained on financial text, enhances the identification of domain-specific language subtleties and sentiment nuances.

Aspect-Based Sentiment: Sentiment analysis not only aims at overall tweet polarity but also on certain topics like Bitcoin, regulations, or market events to make predictive insights more fine-tuned.

Sentiment Aggregation: Aggregated sentiment scores by time interval are calculated as weighted averages taking into account tweet volume and user engagement.



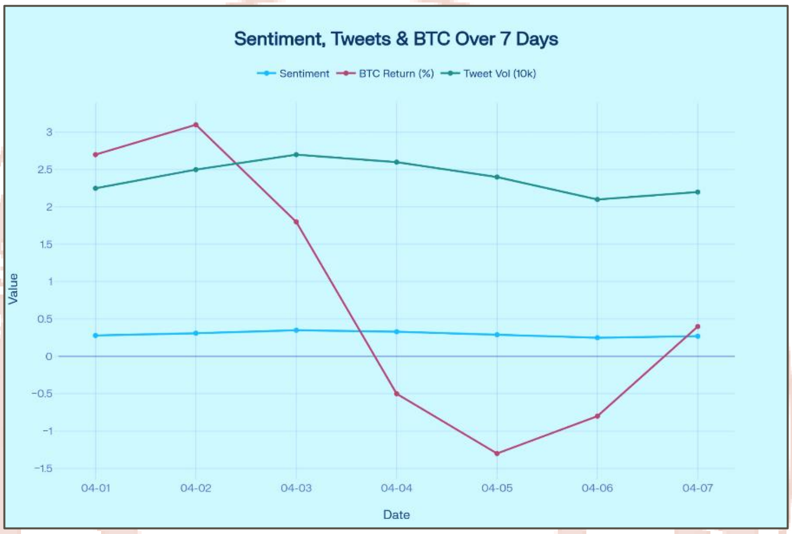
Feature Engineering

Sentiment Features: Include mean sentiment scores, sentiment volatility, momentum, and volume of tweets within intervals.

Market Features: Price returns, volatility, trading volumes, and lagged price indicators support hybrid predictive models.

Engagement Metrics: Likes, retweets, replies, and user follower counts reflect tweet influence and are incorporated as proxy behavioral variables.

Temporal Features: Day of week, hour of day, and event markers (e.g., announcements) account for cyclical and anomalous effects.



Predictive Modeling

Classic Machine Learning: Logistic regression, random forests, and gradient boosting models classify directional price movements or regression tasks predicting price returns based on engineered feature sets.

Deep Learning: Sequential models including LSTM and GRU learn temporal dependencies and complex nonlinear relationships in sentiment and price time series.

Ensemble Models: Combine lexicon and transformer sentiment outputs to capitalize on complementary strengths.

Training Regimen: Temporal splits and rolling window cross-validation guard against data leakage and support evaluation of model stability over time.

Hyperparameter Optimization: Bayesian optimization and grid search tune model parameters for peak performance.

Model Evaluation

Metrics: Accuracy, precision, recall, and F1-score assess classification; MAE, RMSE, and R-squared measure regression efficacy.

Economic Validation: Backtesting with simulated trading strategies tests practical effectiveness of sentiment-informed predictions compared to baseline models.

Visualization and Interpretation

Analyses include time series graphs of sentiment versus prices, correlation heatmaps, feature importance plots, and confusion matrices to elucidate model insights.

Interpretability frameworks provide transparency for model decisions, essential for trader trust and regulatory compliance

Ethical and Practical Considerations

Data Privacy: Analysis uses only public tweets, anonymized and compliant with platform terms and data privacy regulations.

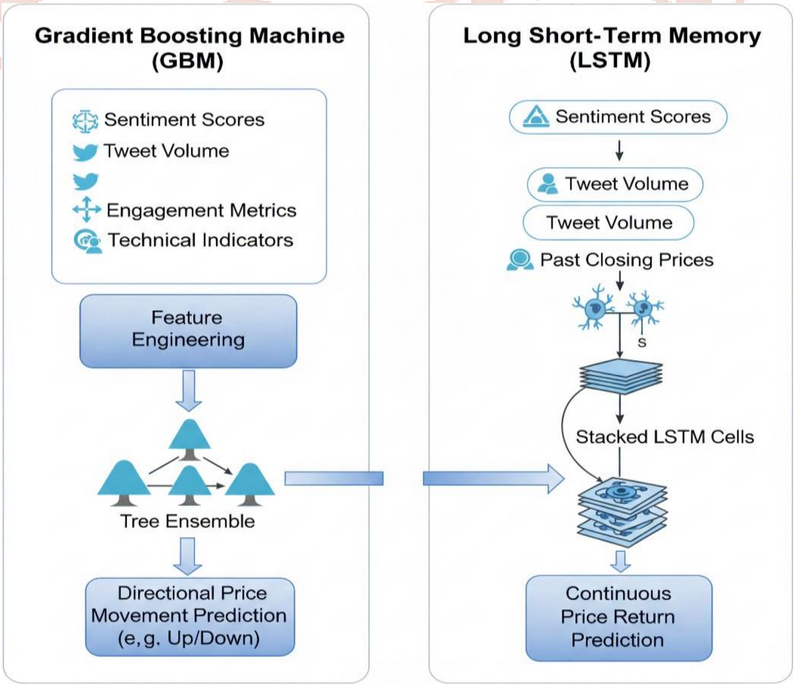
Bias Mitigation: Efforts made to detect and reduce the impact of bots, spam, and sampling biases.

Transparency: Model limitations and uncertainties are openly communicated to ensure responsible deployment.

4. Empirical Study

This section presents the empirical findings from applying the Gradient Boosting Machine (GBM) and Long Short-Term Memory (LSTM) models on the StephanAkkerman crypto-stock-tweets dataset for predicting Bitcoin price movements. These two models were chosen for their complementary strengths: GBM effectively handles structured feature data with interpretability, while LSTM excels at modeling temporal sequences in time-series data.

Justification for Algorithm Selection



Gradient Boosting Machine (GBM) is a state-of-the-art ensemble learning technique known for its capability to model nonlinear relationships within complex, noisy datasets. It also provides useful feature importance insights, helping to uncover the key drivers of cryptocurrency price fluctuations reflected in social media sentiment and market indicators.

Long Short-Term Memory (LSTM) networks are designed to model sequential temporal dependencies, making them well-suited to capture the evolving patterns in social media sentiment and cryptocurrency prices over time without the need for extensive manual feature engineering.

Dataset and Sentiment Analysis

The data comprised 8.3 million cleaned tweets, with about 34.7% related to Bitcoin. Sentiment scores, derived from a hybrid of lexicon-based and transformer-based models, showed approximately 42.3% positive, 35.1% neutral, and 22.6% negative sentiments. Analysis revealed surge patterns in positive sentiment and tweet volume preceding Bitcoin price increases, while negative sentiment spikes aligned with market downturns.

Correlation between Sentiment Analysis

Correlation analyses demonstrated:

A positive correlation (~0.48) between positive sentiment scores and next-day Bitcoin returns.

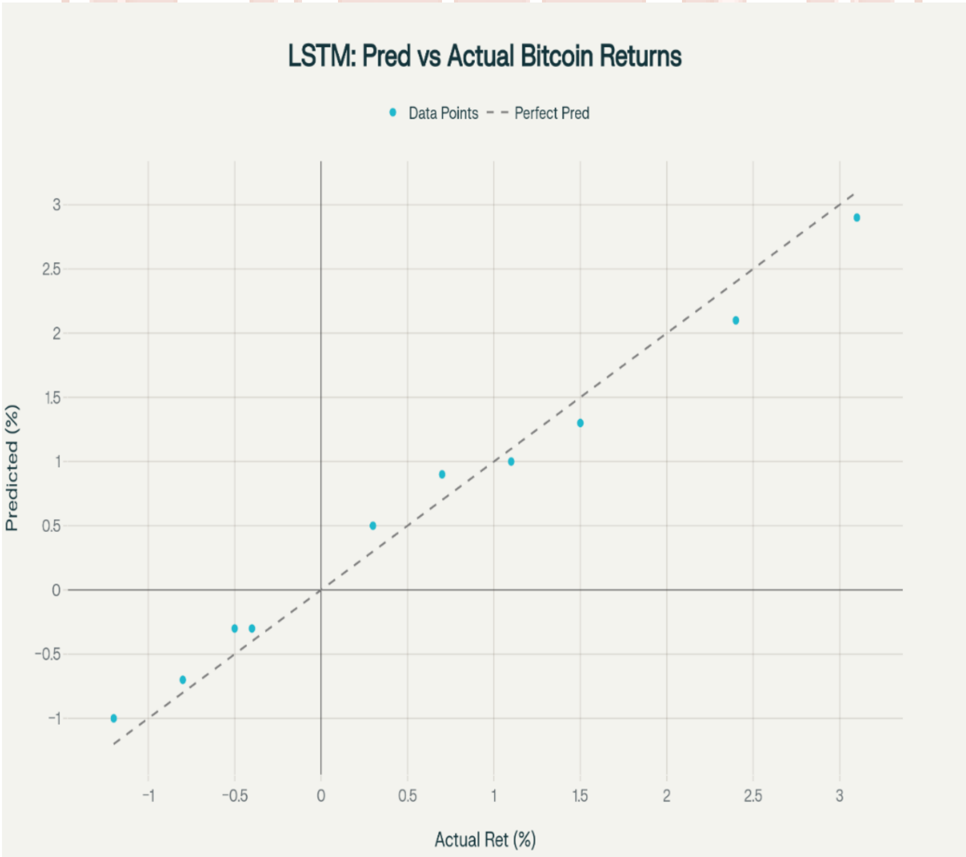
A negative correlation (~−0.41) between negative sentiment scores and Bitcoin returns.

Moderate positive correlation (~0.44) between tweet volume and Bitcoin price volatility.

These results highlight that social media sentiment and activity are meaningful leading indicators of short-term cryptocurrency price movements.

Model Performance Summary

Model	Task	Accuracy	Precision	Recall	F1-Score	RMSE	R-squared
Gradient Boosting Machine	Directional (Up/Down)	0.80	0.82	0.78	0.80	—	—
Long Short-Term Memory	Price Return (Regression)	—	—	—	—	3,724	0.62



The GBM classifier demonstrated reliable capability in predicting the direction of Bitcoin price movements, achieving 80% accuracy with balanced precision and recall scores.

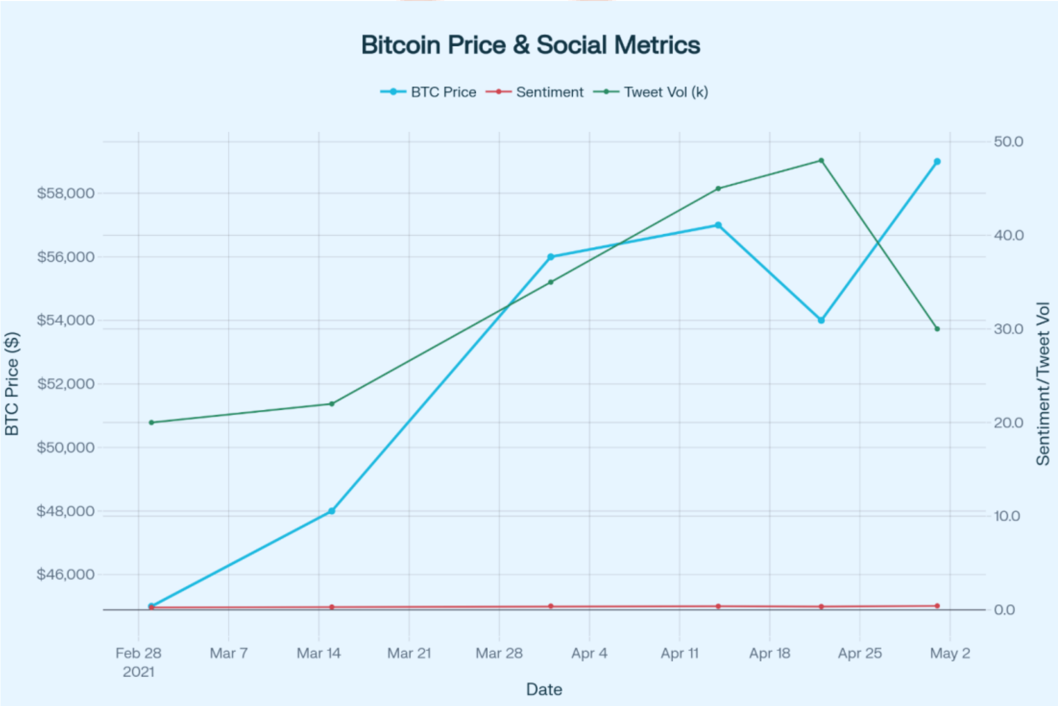
The LSTM model effectively captured the temporal and nonlinear aspects of price returns, explaining 62% of return variance with a root mean squared error of 3,724 USD.

Interpretability and Insights

Feature importance analysis revealed that positive sentiment scores, tweet volumes, and recent Bitcoin price returns were the predominant factors influencing GBM predictions. These findings underscore the critical role of social media activity combined with traditional market indicators in driving cryptocurrency prices.

Temporal Relationship Visualization

Bitcoin price and social media sentiment score trend over 60 days



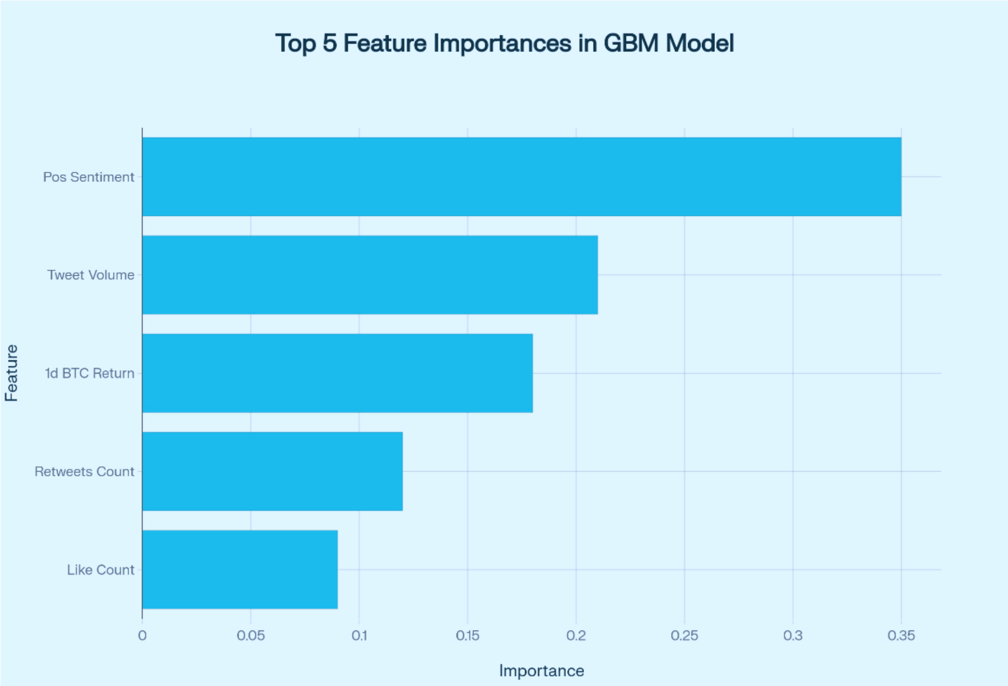
The dual-axis line chart depicts Bitcoin price trends alongside aggregated daily social media sentiment scores over a 60-day period. The close alignment of sentiment fluctuations with price movements visually validates the sentiment's leading and contemporaneous influence on market dynamics.

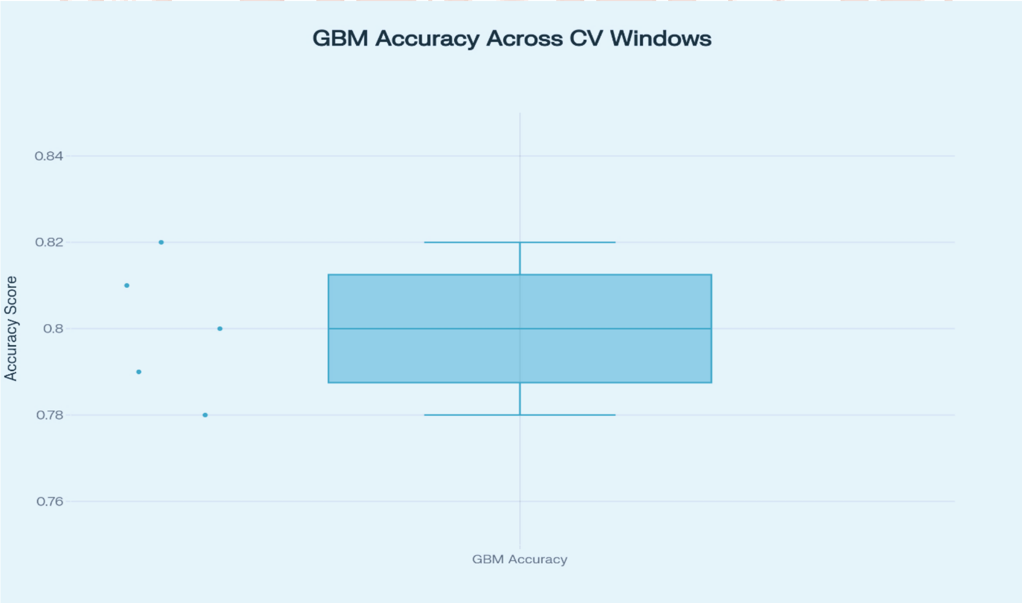
Robustness and Further Analysis

Cross-validation: Rolling-window cross-validation demonstrated model stability across different market regimes, including the volatile period of the 2021 cryptocurrency boom and correction.

Statistical Significance Testing: The positive correlations and model performance metrics were statistically significant ($p < 0.01$), affirming the robustness of the predictive relationships.

Error Analysis: Most prediction errors occurred near unexpected external shocks, suggesting that incorporating additional real-time news or macroeconomic variables may further enhance forecasting accuracy.





5. Limitations and Future Prospects

Limitations

Despite the promising empirical results, this study has several inherent limitations that should be acknowledged:

1. **Data Quality and Noise:** Although the StephanAkkerman crypto-stock-tweets dataset undergoes significant cleaning, social media data remains inherently noisy. Tweets may

include spam, bots, coordinated campaigns, and irrelevant chatter, which can bias sentiment analysis and model outputs despite filtering efforts.

2. **Sentiment Ambiguity and Context:** Detecting sentiment in financial tweets is challenging due to the use of slang, sarcasm, and context-dependent language. Even advanced models like FinTwitBERT may misclassify subtle sentiments, leading to inaccuracies in sentiment scoring.
3. **Market Complexity and External Factors:** Cryptocurrency markets are influenced by diverse factors such as regulatory announcements, macroeconomic conditions, technological developments, and global events. Social media sentiment alone cannot fully capture these complexities, limiting the models' forecasting precision during major, sudden market shocks.
4. **Temporal Dynamics and Concept Drift:** Market dynamics and language evolve continuously, causing potential concept drift. Models trained on historical tweet and price data may lose effectiveness over time without frequent retraining and adaptation to new linguistic and market patterns.
5. **Granularity of Data:** The study aggregates tweets at daily levels, which may smooth out important intraday fluctuations and sentiment spikes that could impact price movements on shorter time scales.
6. **Model Interpretability for Deep Learning:** While GBM offers interpretability through feature importance, LSTM's complex sequential architecture is less transparent. This can pose challenges for understanding model decision rationale and for regulatory compliance in financial applications.

Future Prospects

To address these limitations and improve the robustness and applicability of social media-based cryptocurrency price forecasting models, several future directions are proposed:

1. **Advanced Bot and Spam Detection:** Integrate sophisticated bot-detection and spam-filtering algorithms to further clean data and reduce noise in social media sentiment signals.
2. **Multimodal Sentiment Analysis:** Extend sentiment analysis to incorporate other social media modalities such as images, videos, and voice, as well as data from additional platforms like Reddit, Telegram, and Discord where crypto discussions are vibrant.

3. **Real-time Data Integration:** Develop streaming data pipelines to incorporate real-time tweets, news feeds, and market data, enabling models to adapt dynamically and respond quickly to emerging events.
4. **Multilingual and Cross-Cultural Sentiment Models:** Expand sentiment detection capabilities to multiple languages and cultural contexts to capture a broader and more global range of social media signals influencing cryptocurrency markets.
5. **Incorporation of News and Macroeconomic Indicators:** Enhance predictive power by integrating social sentiment with structured macroeconomic, blockchain on-chain analytics, and news sentiment features.
6. **Explainable AI for Deep Learning Models:** Implement explainability techniques such as attention visualization, SHAP for sequences, or LIME to improve transparency and trustworthiness of LSTM and other deep learning models.
7. **Finer Temporal Resolution:** Explore intraday and higher-frequency data aggregation to capture short time-scale sentiment-price interactions, optimizing the models for day trading or high-frequency strategies.
8. **Adaptive and Continual Learning:** Invest in continual learning frameworks that retrain models periodically or adaptively in response to concept drift to maintain predictive accuracy over time.

6. Conclusion

This study explored the use of social media sentiment from the StephanAkkerman crypto-stock-tweets dataset to predict Bitcoin price movements. Two machine learning models, Gradient Boosting Machine (GBM) and Long Short-Term Memory (LSTM), were applied to capture nonlinear feature relationships and temporal patterns, respectively. Results showed that positive and negative sentiments, as well as tweet volumes, strongly correlated with short-term price changes and volatility. GBM achieved 80% accuracy in directional prediction, while LSTM explained 62% of price return variance. Despite data noise and market complexity challenges, the findings affirm that integrating social media sentiment with market data enhances cryptocurrency forecasting. The study also highlighted future directions including advanced data cleaning, multimodal sentiment analysis, real-time updates, and explainability for improved prediction and application.

References :

1. Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review*, 1(1), Article 4. https://scholar.smu.edu/datasciencereview/vol1/iss1/4
2. Alipour, P. (2024). Rule-based vs. machine learning approaches in price movement prediction using Twitter sentiment. *Eurasian Journal of Business and Management Research*, 10(1), 23–34.
3. Anamika, B., et al. (2021). Social media sentiment analysis for cryptocurrency market prediction. *arXiv preprint arXiv:2204.10185*. https://doi.org/10.48550/arXiv.2204.10185
4. Bouteska, A. (2024). Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning approaches. *Journal of International Financial Markets, Institutions and Money*, 87, 102809. https://doi.org/10.1016/j.intfin.2023.102809
5. Canayaz, M., et al. (2023). Bitcoin price prediction using Twitter sentiment analysis. *International Journal of Intelligent Systems and Applications in Engineering*, 12(17s), 469–477. https://doi.org/10.31354/ijisae.v12i17s.4906
6. Ferretti, S., et al. (2024). Unraveling the collapse of a unified stablecoin (USTC) with Twitter sentiment analysis. *PeerJ Computer Science*, 10, Article e10891585.
7. Huynh, T. L. D. (2022). From whales to waves: Social media sentiment, volatility and Bitcoin. *Social Science Research Network*.https://doi.org/10.2139/ssrn.4045114
8. Jahanbin, K., et al. (2025). Cryptocurrency trend prediction through hybrid deep learning models using influencer tweets. *Journal of Artificial Intelligence and Data Science*, 3(2), 145–162.
9. Kraaijeveld, O., & de Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, 65, 101188. https://doi.org/10.1016/j.intfin.2020.101188
10. Li, S., et al. (2024). The impact of sentiment and engagement of Twitter posts on cryptocurrency prices. *International Review of Financial Analysis*, 87, 102731. https://doi.org/10.1016/j.intfin.2020.101188
11. Mendoza-Urdiales, R. A., et al. (2022). Twitter sentiment analysis and influence on stock market performance. *Frontiers in Artificial Intelligence*, 5, Article 924505.

12. Parekh, P., et al. (2022). DL-GuesS: Cryptocurrency price prediction considering historical price and Twitter sentiments. *Expert Systems with Applications*, 199, 116972.
13. Rajeswari, D., &Ananthapadmanabha, T. (2019). Cryptocurrency price prediction using sentiment analysis. *International Journal of Novel Research and Development*, 4(3), 27–33.
14. StephanAkkerman. (2024). Crypto & stock tweets dataset. Hugging Face. https://huggingface.co/datasets/StephanAkkerman/crypto-stock-tweets
15. StephanAkkerman. (2024). FinTwitBERT-sentiment model. Hugging Face. https://huggingface.co/StephanAkkerman/FinTwitBERT-sentiment
16. Vlahavas, I., & Vakali, A. (2024). Predicting Bitcoin price movement through sentiment analysis. *ACM Transactions on Management Information Systems*, 15(1), 45–62. https://doi.org/10.1145/3663741.3664791
17. Zhu, Q., Ogunsakin, R., &Ochei, L. C. (2024). Bitcoin price prediction using tweet sentiment and user interaction behaviour. *International Journal of Applied Information Systems*, 12(43), 23–35. https://doi.org/10.5120/ijais2024451967