Indonesian Journal of Data and Science

Volume 6 Issue 2
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i3.342

Research Article

Sentiment Analysis of Public Opinion on Pi Network on Reddit Using FinBERT

Sindy Indira Wiguna [1,*]; Adhitia Erfina [2]; Cecep Warman [3]
[1] Nusa Putra University, Sukabumi, Indonesia, sindy.indira_si21@nusaputra.ac.id
[2] Nusa Putra University, Sukabumi, Indonesia, adhitia.erfina@nusaputra.ac.id
[3] Nusa Putra University, Sukabumi, Indonesia, cecep.warman@nusaputra.ac.id

Correspondence should be addressed to Sindy Indira Wiguna; sindy.indira_si21@nusaputra.ac.id

Received 25 September 2025; Accepted 15 December 2025; Published 31 December 2025

© Authors 2025. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes).
License: https://creativecommons.org/licenses/by-nc/4.0/ — Published by Indonesian Journal of Data and Science.

Abstract:
The rapid growth of blockchain technology has led to the emergence of new cryptocurrencies, including Pi Network, which emphasizes accessibility through mobile-based mining. This study aims to answer the research question of whether FinBERT, a financial domain-specific transformer model, can effectively classify public sentiment in informal Reddit discussions related to Pi Network. FinBERT was first evaluated on a labeled financial sentiment dataset to assess its performance in a structured financial context before being applied to Reddit data. Model performance was measured using accuracy, precision, recall, and F1-score. After validation, the model was used to analyze one thousand twenty Reddit comments discussing Pi Network. Text preprocessing included cleaning, case folding, tokenization, stopword removal, stemming, and sequence standardization. The evaluation results show that FinBERT achieved an accuracy of eighty-five point ninety-eight percent on the financial validation dataset, with strong precision and recall across sentiment classes. When applied to Reddit comments, neutral sentiment was the most dominant, followed by positive and negative sentiments. Pi Network was selected as the case study because, unlike more established cryptocurrencies, it is still in an early stage of development and relies heavily on community participation, making public opinion particularly important for understanding its adoption and credibility.

Keywords: : FinBERT, Sentiment Analysis, Reddit, Cryptocurrency, Pi Network

1. Introduction

Blockchain technology has transformed the global financial ecosystem by enabling decentralized transactions and peer-to-peer interactions without relying on intermediaries [1]. Cryptocurrencies such as Bitcoin and Ethereum demonstrate how distributed ledger systems can enhance financial inclusion and create new economic opportunities [2]. Pi Network has emerged as an alternative cryptocurrency that emphasizes accessibility through mobile-based mining, aiming to reach users who are often excluded from conventional crypto adoption [3].

In parallel, social media platforms have become essential venues for discussions about cryptocurrencies. Reddit, in particular, hosts dynamic communities where optimism, skepticism, and neutrality coexist in ongoing debates about blockchain projects [4]. Previous studies indicate that sentiment expressed on online platforms significantly influences cryptocurrency adoption, market volatility, and investor confidence [5], [6]. Consequently, analyzing sentiment from user-generated discussions provides valuable insights into how public perception shapes the adoption trajectory of blockchain initiatives.

Sentiment analysis has evolved from lexicon-based and traditional machine learning approaches to deep learning and transformer-based models [7]. The introduction of BERT enabled contextualized representations and transfer

learning, significantly improving sentiment classification performance [8]. FinBERT, a domain-specific variant of BERT trained on financial corpora, has demonstrated superior performance in sentiment analysis tasks involving structured financial texts such as analyst reports and financial news [9], [10], [11].

However, most existing FinBERT studies remain focused on formal and structured financial documents. Empirical investigations addressing informal, user-generated discussions on Reddit are still limited, particularly in the context of emerging cryptocurrencies. Prior cryptocurrency sentiment studies using FinBERT or other NLP models predominantly analyze well-established assets, such as Bitcoin and Ethereum, leaving projects like Pi Network underexplored. This limitation is important because Reddit discussions are linguistically diverse, informal, and often speculative, posing distinct challenges for domain-specific language models.

Therefore, this study aims to evaluate whether FinBERT, fine-tuned on a labeled financial dataset, can effectively analyze sentiment in informal Reddit discussions related to Pi Network. The objectives and contributions of this research are aligned with this gap by assessing FinBERT's adaptability to noisy social media data and by providing empirical evidence of public sentiment toward an emerging cryptocurrency project.

2. Method

Study adopted a systematic workflow to analyze public sentiment toward Pi Network on Reddit using FinBERT. The methodological design consisted of dataset preparation, preprocessing, model fine-tuning, evaluation, and inference on Reddit comments. This approach was selected to evaluate FinBERT's performance on structured financial data before assessing its adaptability to informal, user-generated discussions.

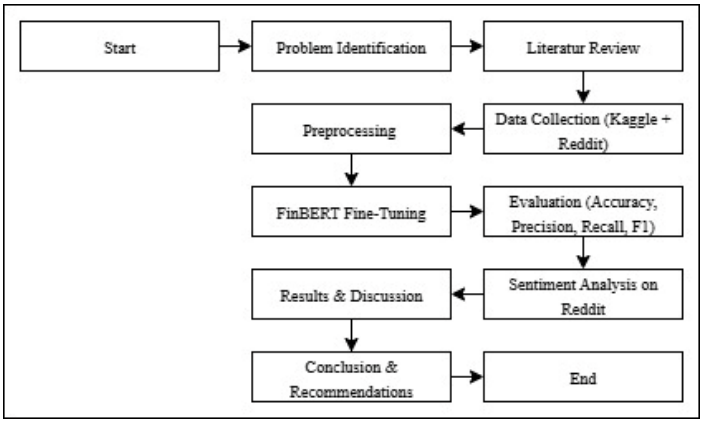

Figure 1. Research Flow

Datasets

Two datasets were employed. The first was a labeled financial sentiment dataset obtained from Kaggle, consisting of four thousand eight hundred forty-seven samples categorized into positive, neutral, and negative sentiments. This dataset was used for fine-tuning and validation. The second dataset consisted of one thousand twenty Reddit comments related to Pi Network, which were collected through web crawling and used exclusively for inference.

Preprocessing

Both datasets underwent preprocessing to reduce noise and improve model robustness. The preprocessing steps included cleaning, case folding, tokenization using the FinBERT tokenizer, stopword removal, stemming, and padding and truncation to a fixed sequence length of one hundred twenty-eight tokens.

Although transformer-based models such as FinBERT rely on contextual embeddings and are generally less sensitive to surface-level text variations, stopword removal and stemming were applied to reduce lexical sparsity and normalize informal expressions commonly found in Reddit comments. This was particularly important given the presence of repetitive function words, slang variations, and morphological inconsistencies in user-generated text. The

preprocessing pipeline was applied consistently across both datasets to maintain comparability between training and inference stages.

Fine-Tuning and Evaluation

FinBERT was fine-tuned on the financial sentiment dataset using an eighty to twenty split for training and validation. A softmax classification layer was added to classify inputs into three sentiment classes: positive, neutral, and negative. Model performance was evaluated using accuracy, precision, recall, F1-score, and a confusion matrix.

Class Imbalance Considerations

The training dataset exhibited a moderate class imbalance, with the neutral class being more dominant than positive and negative classes. No explicit rebalancing techniques, such as oversampling or class-weight adjustment, were applied during training. As a result, the model may have developed a bias toward predicting neutral sentiment, which is reflected in both the validation results and the sentiment distribution observed in Reddit inference. This limitation is acknowledged as a potential factor influencing classification outcomes and highlights the importance of addressing class imbalance in future research.

3. Result and Discussion

Model Evaluation on Financial Dataset

The fine-tuned FinBERT model demonstrated strong performance on the financial validation dataset, achieving an accuracy of eighty-five point ninety-eight percent. This performance is consistent with previous studies showing that FinBERT outperforms generic transformer models in financial sentiment classification due to its domain-specific pre-training [2], [6], [16]. The neutral class achieved the highest precision and recall, indicating the model's effectiveness in identifying non-polarized financial statements, which are common in financial discourse.

Misclassifications were mainly observed between the positive and neutral classes, a pattern also reported in earlier sentiment analysis studies, where weakly polarized expressions often blur class boundaries [14], [19].

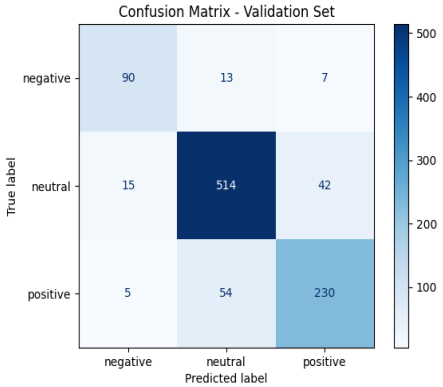

Figure 2. Confusion Matrix of FinBERT on Kaggle validation dataset.

Sentiment Distribution on Reddit Data

After validation, the fine-tuned FinBERT model was applied to one thousand twenty Reddit comments related to Pi Network. The resulting sentiment distribution indicates that neutral sentiment is the most dominant category, followed by positive and negative sentiments, as illustrated in **Figure 3**.

```
Akurasi Model: 0.8598

Classification Report:
              precision    recall  f1-score   support

    negative       0.82      0.82      0.82       110
     neutral       0.88      0.90      0.89       571
    positive       0.82      0.80      0.81       289

    accuracy                           0.86       970
   macro avg       0.84      0.84      0.84       970
weighted avg       0.86      0.86      0.86       970
```

Figure 3. Sentiment Distribution of Reddit Comments

The predominance of neutral sentiment suggests that Reddit users tend to engage in descriptive and informational discussions rather than expressing strong emotional polarity. This pattern is consistent with findings from previous cryptocurrency sentiment studies on Reddit and Twitter, which report that neutral sentiment often dominates discussions related to blockchain projects and digital assets [17], [18], [21].

Similar results have been observed in studies focusing on established cryptocurrencies such as Bitcoin and Ethereum, where users frequently share technical updates, market speculation, and general information without explicit positive or negative framing [17], [18]. Therefore, the dominance of neutral sentiment in Pi Network discussions reflects a broader characteristic of cryptocurrency-related online discourse rather than an anomaly specific to this dataset.

Interpretation of Sentiment Polarity (Positive, Neutral, and Negative)

Positive sentiment in the Reddit discussions reflects optimism regarding Pi Network's potential adoption, perceived future value, and accessibility through mobile-based mining. Similar expressions of optimism have been reported in previous studies on emerging or speculative cryptocurrencies, where community-driven expectations play a significant role in sustaining engagement during early development stages [5], [11], [17]. This interpretation is supported by the positive sentiment word cloud, which highlights frequently occurring terms such as "will," "pi," "price," "buy," "good," and "going," indicating expectations of future growth and investment potential.

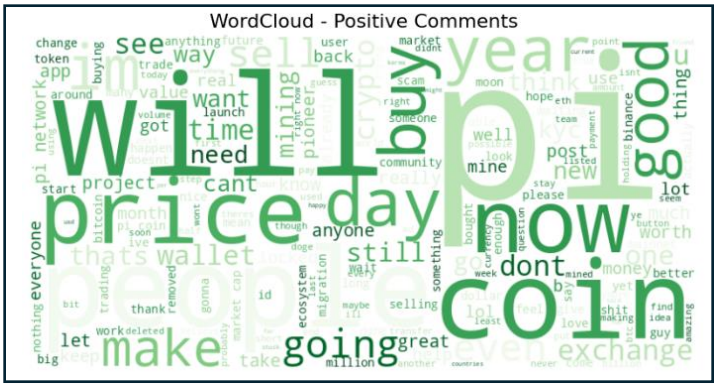

Figure 4. Positive Sentiment Word Cloud

Neutral sentiment, which dominates the discussion, primarily represents descriptive and informational exchanges among users. These comments often focus on technical aspects such as mining processes, account verification, and general project updates without strong emotional framing. Prior research on cryptocurrency-related social media discourse has similarly identified neutral sentiment as a common category, reflecting information sharing rather than opinionated debate [17], [18], [21]. This interpretation is supported by the neutral sentiment word cloud, which highlights frequently occurring terms such as "mine," "kyc," "sell," and "people," indicating discussions centered on mining activities, verification processes, and general user experiences.

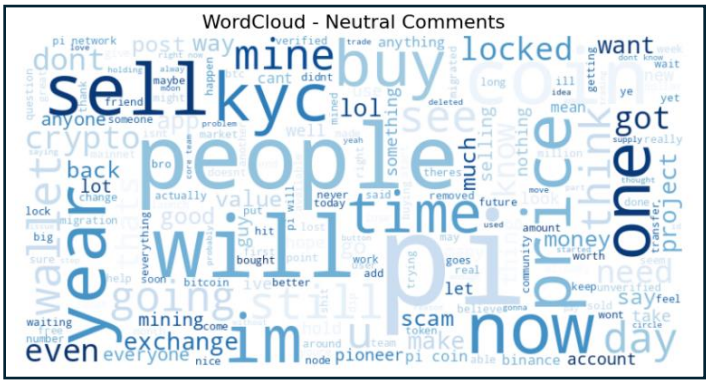

Figure 5. Neutral Sentiment Word Cloud

Negative sentiment highlights skepticism and concerns related to project credibility, transparency, and long-term viability. Studies examining cryptocurrency communities consistently report distrust and skepticism toward newer projects that have not yet achieved full market integration or regulatory clarity [5], [18]. The coexistence of optimism, neutrality, and skepticism observed in this study suggests a cautious but engaged community response, which is characteristic of emerging blockchain initiatives. This interpretation is supported by the negative sentiment word cloud, which frequently includes terms such as “scam,” “sell,” “coin,” and “buy,” reflecting doubts and critical attitudes toward Pi Network’s legitimacy and future value.

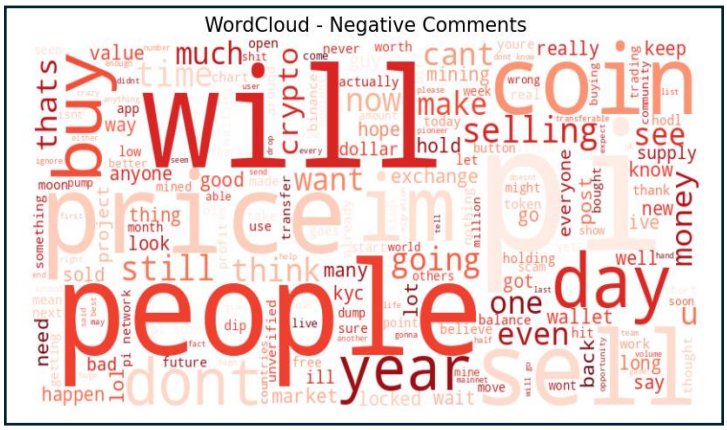

Figure 6. Negative Sentiment Word Cloud.

Discussion

This study analyzed public sentiment toward Pi Network on Reddit by fine-tuning and applying FinBERT, a financial domain-specific transformer model. The results demonstrate that FinBERT performs well on structured financial validation data and can be effectively extended to informal, user-generated Reddit discussions. The sentiment distribution indicates that neutral sentiment dominates Pi Network-related discussions, accompanied by positive expressions of optimism and negative expressions of skepticism. These findings provide empirical evidence of public perception toward an emerging cryptocurrency project and confirm FinBERT’s adaptability beyond formal financial texts.

One key limitation of this study is the domain mismatch between the training data and the inference data. FinBERT was fine-tuned on a labeled financial dataset, while Reddit comments exhibit informal language, slang, and stylistic variability. This domain difference may contribute to prediction bias, particularly toward neutral sentiment, and should be considered when interpreting the results.

Future research may address this limitation by incorporating domain-adaptive pre-training on social media corpora or by comparing FinBERT with transformer models specifically adapted to social media text. Additional improvements may include handling class imbalance through data rebalancing techniques and expanding the dataset

to include multiple social platforms. These directions may further enhance sentiment classification performance and improve the robustness of sentiment analysis for emerging cryptocurrency discussions.

4. Conclusion

This study investigated public sentiment toward Pi Network by applying the FinBERT model to Reddit discussions. The results demonstrate that FinBERT, when fine-tuned on a labeled financial dataset, performs effectively in capturing sentiment patterns within informal social media text. The sentiment distribution revealed a dominance of neutral opinions, reflecting the exploratory and informational nature of many Reddit discussions, while positive and negative sentiments highlighted contrasting expectations regarding Pi Network's potential value and long-term viability.

The main contribution of this research lies in extending the application of FinBERT beyond formal financial documents to user-generated content from Reddit, thereby providing insights into sentiment dynamics surrounding an emerging cryptocurrency. However, one limitation of this study is the domain mismatch between the financial texts used for model training and the informal language found in Reddit comments, which may influence classification performance, particularly for nuanced or ambiguous expressions.

Future research may address this limitation by comparing FinBERT with transformer models specifically adapted for social media data or by incorporating domain-adaptive pretraining techniques. Additionally, expanding the analysis to multiple social media platforms and conducting temporal sentiment analysis could offer a more comprehensive understanding of public opinion dynamics in cryptocurrency communities.

ACKNOWLEDGMENT

The author would like to thank all parties who contributed to the completion of this research, particularly those who provided constructive feedback and support during the data collection and analysis process. No external funding was received for this study.

References

[1] A. H. Huang, H. Wang, and Y. Yang, "FinBERT: A large language model for extracting information from financial text," *Contemp. Account. Res.*, vol. 40, no. 2, pp. 806–841, 2023, doi: 10.1111/1911-3846.12832.

[2] Y. Yang, M. C. S. Uy, and A. H. Huang, "FinBERT: A pretrained language model for financial communications," arXiv preprint.

[3] S. Gururangan et al., "Don't stop pretraining: Adapt language models to domains and tasks," in *Proc. ACL*, 2020, pp. 8342–8360, doi: 10.18653/v1/2020.acl-main.740.

[4] S. Kumar, P. P. Roy, D. P. Dogra, and B.-G. Kim, "A comprehensive review on sentiment analysis: Tasks, approaches and applications," *Inf. Fusion*, vol. 89, pp. 384–401, 2023, doi: 10.1016/j.inffus.2022.09.015.

[5] E. Mnif, A. Jarboui, and K. Mouakhar, "How the cryptocurrency market has performed during COVID-19?" *Finance Res. Lett.*, vol. 36, p. 101647, 2020, doi: 10.1016/j.frl.2020.101647.

[6] M. Cary, "Herding and investor sentiment after the cryptocurrency crash: Evidence from Twitter and natural language processing," *Financial Innovation*, vol. 10, p. 63, 2024, doi: 10.1186/s40854-024-00663-x.

[7] M. Kraus, S. Feuerriegel, and A. Oztekin, "Wisdom of the crowd signals: Predictive power of social media trading signals for cryptocurrencies," *Electron. Mark.*, 2025, doi: 10.1007/s12525-025-00815-6.

[8] Y. Kim, J. Kim, and A. Oh, "Explaining cryptocurrency price trends using social media data," in *Proc. Web Conf.*, 2024, doi: 10.1145/3654823.3654866.

[9] R. Issalh, A. Gupta, and M. Gupta, "Pi Network: A revolution," *Sci. J. Metaverse Blockchain Technol.*, vol. 1, no. 1, pp. 18–27, 2023, doi: 10.36676/sjmbt.v1i1.03.

[10] K. M. S. Samin and B. K. Deshmukhya, "A brief analysis on predicted future value of Pi Network on the basis of Bitcoin," *Glob. Mainstream J. Innov. Eng. Emerg. Technol.*, 2022.

[11] S. A. Dauda, I. Y. Aliyu, and J. Ibrahim, "Social media and cryptocurrency adoption: Evidence from Pi Network communities," 2021.

[12] Y. Tang et al., "FinEntity: Entity-level sentiment classification for financial texts," arXiv preprint.

[13] Z. Han et al., "Parameter-efficient fine-tuning for large models: A comprehensive survey," arXiv preprint.

[14] S. Sathyanarayanan, "Confusion matrix-based performance evaluation metrics," *Afr. J. Biomed. Res.*, vol. 27, no. 4s, pp. 4023–4031, 2024, doi: 10.53555/ajbr.v27i4s.4345.

[15] M. Saputra and S. Wahyuni, "Sentiment analysis of digital banking application reviews using support vector machine," *INFOTECH J.*, vol. 10, no. 2, pp. 327–332, 2024, doi: 10.31949/infotech.v10i2.11801.

[16] A. Sinaga and S. P. Nainggolan, "Comparison of stemming algorithms for Indonesian text processing," *Sebatik*, vol. 27, no. 1, pp. 63–69, 2023, doi: 10.46984/sebatik.v27i1.2072.

[17] T. M. Fahrudin et al., "Word cloud visualization for sentiment interpretation in social media analysis," *SKANIKA*, vol. 5, no. 2, pp. 190–202, 2022.

[18] T. Jamaluddin, M. A. Bijaksana, and I. Asror, "Comparison of SentencePiece tokenization algorithms for text processing," 2020.

[19] R. Zhang, R. Xue, and L. Liu, "Security and privacy issues in blockchain applications," *IEEE Trans. Serv. Comput.*, vol. 15, no. 6, pp. 3668–3686, 2022, doi: 10.1109/TSC.2021.3085913.

[20] Z. Hasan et al., "Blockchain and cryptocurrency regulation challenges," *Birokrasi*, vol. 2, no. 2, pp. 55–69, 2024, doi: 10.55606/birokrasi.v2i2.1163.

[21] A. Kumar and T. Sebastian, "Sentiment analysis of online financial discussions: A survey," *Expert Syst. Appl.*, 2021.

[22] F. Wilyani, Q. N. Arif, and F. Aslimar, "Introduction to Python programming using Google Colaboratory," J. *Pelayanan dan Pengabdian Masy. Indones.*, vol. 3, no. 1, pp. 8–14, 2024, doi: 10.55606/jppmi.v3i1.1087.

[23] N. Anggraini et al., "Text-based prediction using deep learning and attention mechanisms," 2025.

[24] B. Arianto, "Blockchain governance and digital currency adoption," *J. Soc. Polit. Gov.*, 2020.

[25] M. Cary, "Investor sentiment and cryptocurrency market behavior," *Financial Innovation*, 2023.