

PROJECT PROPOSAL

PROJECT TITLE	Evidence-enhanced disinformation detection for Crypto News
TEAM	1. Priyam Dinesh Shah

I. PROBLEM STATEMENT

The crypto market space is 1.08T and is forecasted to grow higher [8]. The space is dynamically changing and too much news is being spread worldwide to influence the market. This project will help users to get a guess estimated idea of whether certain news related to crypto can be relied on or not. Disinformation is a major problem in the world today, and with the rise of cryptocurrencies, it has become easier for people to spread false information and manipulate the market. Detecting disinformation is a challenging task, especially in the crypto space where there is a lot of noise and misinformation. This project aims to develop and compare machine-learning algorithms that can detect disinformation in crypto news articles by leveraging evidence from multiple sources. The project can then be extended with additional modifications for different news categories.

II. BENEFIT AND OUTCOME

I plan to deliver an ipynb file which will ask for a minimum worded news article from the user. A training dataset and model file if required by the algorithm. The results of the ML-based algorithm tentatively would be a truth value on grayscale (0-100%), a related evidence list supporting the value, and comparison graphs from different approaches. I think this project will benefit me in applying my machine learning subject knowledge to real-world problem-solving. My project is unique as there are not many researchers who have tweaked and applied existing ML methods to crypto news specifically.

III. LITERATURE REVIEW

Disinformation detection has been an active area of research in recent years, and several approaches have been proposed to tackle this problem. These approaches include natural language processing (NLP) techniques, graph-based methods, social network analysis, and machine learning methods.

NLP techniques involve analyzing the text of a document to extract relevant features, such as sentiment, topics, and entities. Graph-based methods use network structures to model the relationships between documents, users, and other entities and can be used to identify patterns of disinformation propagation. Social network analysis focuses on identifying influential users and their interactions, which can be used to identify potential sources of disinformation.

However, these approaches often rely on a single source of evidence and may not take into account the complex interplay between different sources. In contrast, our proposed approach aims to leverage evidence from multiple sources, including social media, news articles, and user comments, to detect disinformation more effectively. By combining evidence from multiple sources, we can create a more comprehensive view of the information space and identify patterns that may be missed by approaches that rely on a single source.

Comparison Study: This is important because the goal of disinformation detection is not only to detect disinformation but also to understand why certain content is classified as disinformation. Evidence urls or sources and Scores provided will allow me to provide clear and understandable explanations of generated results, which can help build trust and confidence in the system.

Overall, my proposed approach differs from existing approaches by leveraging evidence from multiple sources and using comparative studies and evidence of derived results.

IV. Preliminary Plan (MILESTONES)

Sr. No	Milestones
1	Data Collection and Preprocessing: Collect and preprocess a large dataset of crypto news articles and their associated comments and social media posts.
2	Feature Extraction: Extract relevant features from the dataset, including sentiment, topics, entities, and author reputation.
3	Model Development: Develop machine learning models to detect disinformation in the dataset, using various techniques such as natural language processing, graph-based methods, and social network analysis.
4	Model Comparison and Evaluation: Compare the performance of different models and evaluate their effectiveness using various metrics such as precision, recall, and F1 score.
5	Interpretation and Explanation of results.
6	Final Report writing
7	Final Project Presentation

V. CITATIONS

1. Hamdi T., Slimi H., Bounhas I., Slimani Y. (2020) A Hybrid Approach for Fake News Detection in Twitter Based on User Features and Graph Embedding. In: Hung D., D'Souza M. (eds) Distributed Computing and Internet Technology. ICDCI 2020. Lecture Notes in Computer Science, vol 11969. Springer, Cham. https://doi.org/10.1007/978-3-030-36987-3_17
2. Bharadwaj, Pranav and Shao, Zongru, Fake News Detection with Semantic Features and Text Mining (July 24, 2019). International Journal on Natural Language Computing (IJNLC) Vol.8, No.3, June 2019, Available at SSRN: <https://ssrn.com/abstract=3425828>.
3. Zhang, Y., Yang, D., He, D., & Huang, J. (2020). GTC: A Graph-based Text Classification Method for Detecting Disinformation. IEEE Transactions on Knowledge and Data Engineering.
4. Liu, Z., Zhang, Y., He, D., & Huang, J. (2019). A graph-based approach for detecting disinformation in social media. Journal of Computer Science and Technology.
5. Arroyo-Fernández, I., Fernández-Vilas, A., & Anido-Rifón, L. E. (2021). Disinformation Detection in Twitter through Topic Modeling and Sentiment Analysis. International Journal of Interactive Multimedia and Artificial Intelligence.
6. Feng, H., Zhang, J., Gao, L., & Zhang, J. (2020). Multi-modal disinformation detection based on transfer learning. IEEE Access.
7. Chen, W., & Wang, Y. (2019). Disinformation Detection on Social Media via Propagation Path based Learning. Proceedings of the 28th ACM International Conference on Information and Knowledge Management.
8. CoinMarketCap. (2023, February 26). Cryptocurrency Prices, Charts And Market Capitalizations. Retrieved from <https://coinmarketcap.com/>