

Peer Review

Review of: "CryptoPulse: Short-Term Cryptocurrency Forecasting with Dual-Prediction and Cross-Correlated Market Indicators"

Cristian Cosentino¹

1. DIMES, University of Calabria, Italy

The CryptoPulse framework represents a promising step forward in short-term cryptocurrency forecasting by integrating macroeconomic indicators, technical features, and market sentiment derived from news articles. However, when evaluated critically and compared with prior research, several areas emerge that require substantial improvement to elevate the model's reliability, generalizability, and interpretability. The most evident limitation concerns the scope of sentiment analysis. Although the authors utilize GPT-3.5 with a think-tank-inspired prompting strategy, the news source is limited to Cointelegraph, which restricts the diversity and reactivity of sentiment signals. In previous work by Cantini, Cosentino et al. ("Harnessing prompt-based large language models for disaster monitoring and automated reporting from social media feedback", Q1), it was shown how feedback from real-time platforms such as Twitter and Reddit can drastically enrich the situational context and allow large language models to generate more nuanced and time-sensitive analyses. Similarly, the multidimensional classification approach proposed in "Multi-dimensional classification on social media data for detailed reporting with large language models" demonstrated the effectiveness of structured, hierarchical classification in improving model interpretability and content segmentation, which could be adopted here to contextualize different sentiment signals based on topic or source reliability.

Furthermore, CryptoPulse does not sufficiently address the sensitivity of sentiment output to variations in prompt design. As shown in the evaluation from "Enhancing cryptocurrency price forecasting by integrating machine learning with social media and market data" (Belcastro et al., Q2), the stability of predictive systems in volatile domains like crypto markets depends heavily on the robustness of auxiliary

inputs like sentiment. Prompt variability can introduce inconsistencies that affect the weighting mechanisms used in model fusion. This issue is echoed in external studies, such as Zhang et al. (2023) “Sentiment analysis in the era of large language models: A reality check”, which demonstrated how even minimal perturbations in prompt templates lead to significant changes in the predicted label distribution. Moreover, as Si et al. (2022) propose in “Prompting GPT-3 to be reliable”, consistency-based calibration techniques or majority-vote prompting can regularize outputs without adding significant computational cost and should be considered in future versions of CryptoPulse.

The fusion mechanism used to combine macroeconomic and technical predictions is another area that warrants reconsideration. Currently, the model applies a sentiment-weighted linear fusion, but such a method lacks flexibility, particularly in scenarios where sentiment may be ambiguous or conflicting. An attention-based or learnable gating mechanism that adjusts fusion weights dynamically would likely outperform a scalar fusion design. This becomes particularly important in cross-market scenarios where the volatility of low-cap cryptocurrencies may not follow macro trends and thus require more emphasis on local dynamics. Additionally, the model should account for causality in the sentiment-price relationship. CryptoPulse implicitly assumes that news sentiment precedes and influences price movement, but this is not always the case; in fact, price changes often influence the sentiment expressed in retrospective news articles. Temporal disentanglement of such bidirectional relationships is necessary to avoid misleading attribution.

Another critical aspect missing from CryptoPulse is a deeper reflection on evaluation metrics. While standard regression metrics like MAE and MSE are useful, they fail to capture the utility of predictions in real-world scenarios such as algorithmic trading or risk management. Metrics like directional accuracy, profit-and-loss simulation, and Sharpe ratio would offer a more comprehensive view. This was discussed in earlier work such as “Harnessing prompt-based LLMs,” where the generated reports were also evaluated based on their actionability and informativeness, not merely their lexical correctness or correlation with ground truth. Finally, from a reproducibility perspective, CryptoPulse could benefit from clearer documentation of the data curation process, especially regarding how sentiment labels are handled in the case of missing news days, which may create time series gaps that distort normalization. Introducing imputation strategies or leveraging models tolerant of uneven sequences would mitigate these problems.

In conclusion, while CryptoPulse introduces a compelling dual-prediction mechanism and a novel use of LLM-based sentiment analysis, it remains bounded by narrow assumptions about data sources,

underdeveloped fusion strategies, and a lack of robustness studies regarding prompt sensitivity. By adopting lessons from existing literature—including both the cited studies and broader LLM evaluation work—it can evolve into a more comprehensive and stable forecasting tool with stronger predictive fidelity and applicability in real-world financial environments.

Declarations

Potential competing interests: No potential competing interests to declare.