Crypto Sentiment Analysis

Liubomyr Bobrovych
CTU–FIT
bobroliu@fit.cvut.cz

12. 12. 2025

1 Introduction

The goal of this project was to design and implement a system capable of collecting, preprocessing, and analyzing **social-media sentiment** related to selected cryptocurrencies. The aim was to capture market mood from public discussions on Reddit and Telegram and present the results through a CLI pipeline and an interactive Streamlit interface. Unlike traditional financial analysis, this project focuses purely on **textual sentiment signals**, offering a lightweight, real-time perspective on community perception.

2 Data

The data originates from Reddit (via the PRAW API) and Telegram (via synchronous and asynchronous Telethon clients). Users choose a cryptocurrency, platform, and time period; the system then gathers posts matching these conditions.

Because real-world text contains URLs, mentions, Markdown fragments, emojis, and inconsistent formatting, preprocessing consists of:

- **soft cleaning**: removing URLs, mentions, and Markdown links while preserving emojis and casing for accurate VADER scoring,
- **hard cleaning**: lowercasing, removing punctuation and non-ASCII symbols for potential ML extensions,
- merging Reddit titles and post bodies into a unified `combined` field,
- preserving metadata such as author/channel, timestamp (UTC ISO), and engagement values (Reddit upvotes, Telegram views).

These steps ensure consistent structure across both platforms and prepare the text for downstream sentiment computation.

3 Methods and Algorithms

The application uses two collectors: **PRAW** for Reddit submissions and **Telethon** for Telegram channels. After collection, posts are normalized, cleaned, and transformed into a unified format.

Sentiment is computed using the standard **VADER** model, supplemented with a manually crafted lexicon of crypto-community slang and domain-specific terms, enabling more accurate scoring of cryptocurrency discussions. Each message receives a compound VADER sentiment score.

To estimate overall sentiment for the selected platform and time window, the system computes a **weighted average**:

- Reddit uses post upvotes as weight,
- Telegram uses view count.

This gives greater influence to posts that reached a wider audience.

The Streamlit interface provides an interactive environment in which users can:

- configure source, coin, and time period,
- inspect individual messages and their sentiment,
- view sentiment distribution and average sentiment per channel,
- explore posting activity over time.

Development followed a **feature-branch Git workflow**, ensuring isolated updates and clean integration of new components.

4 Results

Overall, the system successfully collects posts, preprocesses them, assigns sentiment scores, and aggregates the results into a clear market mood indicator. Across various tests, several general patterns appeared:

- some Telegram channels consistently show strong positive or negative bias,
- sentiment fluctuates significantly in short periods, reflecting speculation-driven behavior in crypto communities,

- VADER performs well even on informal text containing emojis and slang.

To examine whether community sentiment aligns with real market behavior, a case study was performed for **ETH sentiment on Reddit during 7–12 December 2025**. The results showed a distinctly positive aggregated sentiment as well as a distribution shifted toward positive values. Comparing these findings with ETH price data from CoinMarketCap for the same period revealed a noticeable upward trend. The community reacted with optimism, matching the market's performance.

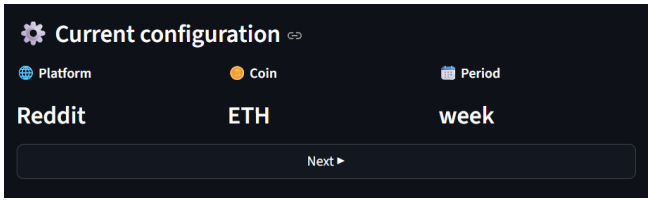

Figure 1: User configuration for ETH sentiment analysis

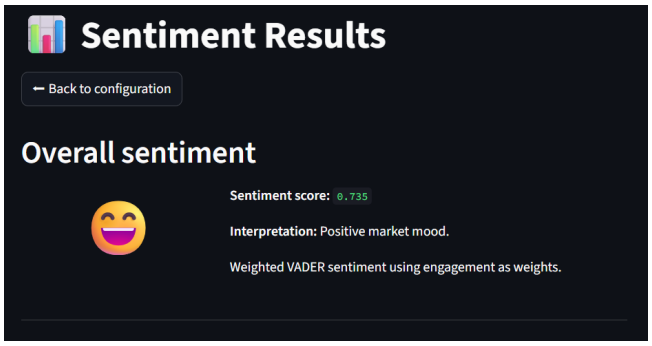

Figure 2: Aggregated sentiment score for ETH (Reddit, 7–12.12.2025)

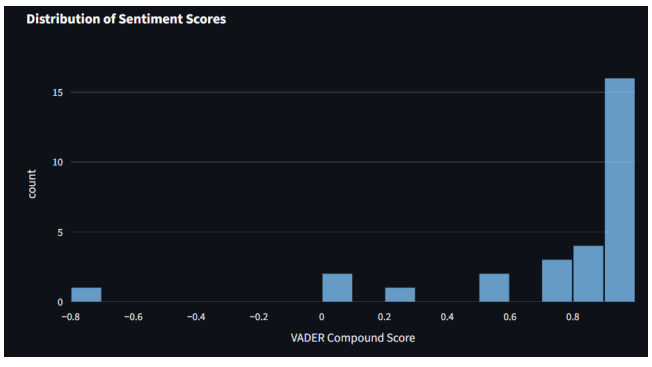

Figure 3: Sentiment distribution for collected ETH posts

These results demonstrate that community sentiment can reflect short-term market dynamics, although sentiment naturally remains influenced by the author's subjective view.

Figure 4: ETH price movement (CoinMarketCap, same period)

5 Conclusion

The project provides a functional and extensible framework for crypto-related sentiment analysis, accessible through both CLI and Streamlit. Its modular design allows further development and experimentation.

Possible extensions include:

- collecting data from more channels and subreddits,
- using emoji reactions, forwards, or comments as additional sentiment weights,
- integrating market data for automated correlation analysis,
- adding crypto news sentiment and embedding-based similarity filtering.

References

- Streamlit Documentation. Available at: `https://docs.streamlit.io/`
- PRAW — Python Reddit API Wrapper Documentation. Available at: `https://praw.readthedocs.io/`
- Telethon Documentation. Available at: `https://docs.telethon.dev/`
- Hutto, C. — *VADER Sentiment Analysis.* GitHub Repository. Available at: `https://github.com/cjhutto/vaderSentiment`