Analysing Sentiment from Financial News Article Titles: A Text-Mining Pipeline with a 2020 Market Shock Case Study

Matz Chan (Match933)
Linköping University
TDDE16 Text Mining
`Match933`

Abstract

This project analysed whether sentiment extracted from financial news article titles shifted around major real-world events and whether it related to realised market performance. I scored a reproducible pool of 3,000 headlines from January–May 2020 using FinBERT, a finance-domain sentiment classifier. I compared sentiment distributions before and after two 2020 events: the World Health Organization (WHO) pandemic declaration and the U.S. Federal Reserve announcement of large-scale asset purchases. I evaluated distribution shifts using a two-sided Mann–Whitney U test and reported Cliff's delta as a non-parametric effect size. Finally, I aggregated daily sentiment and tested its correlation with S&P 500 daily returns as an exploratory validation signal. The results showed statistically significant sentiment shifts around both events and a positive correlation between daily sentiment and daily market returns in the same period.

1 Introduction

Financial markets react rapidly to information, and headlines are often the earliest compressed form of that information. In this project, I built and evaluated a text-mining pipeline for sentiment analysis on financial news article titles. I used early 2020 as a concrete case study because it contained multiple widely discussed market-moving events and large price swings, but the pipeline was designed to generalise to other time windows and event sets.

I asked two research questions:

- **RQ1:** Did headline sentiment differ before and after major events?
- **RQ2:** Was daily aggregated headline sentiment correlated with daily S&P 500 returns?

2 Related Work

Transformer-based models are widely used for sentiment classification. BERT introduced pretraining objectives that improved performance across many NLP tasks (Devlin et al., 2019). FinBERT adapted BERT to financial text and is commonly used for finance-specific sentiment inference (ProsusAI, 2025). As a baseline, I also computed VADER sentiment scores, a lexicon- and rule-based method often used for short texts (Hutto and Gilbert, 2014).

3 Data

3.1 News headlines

I used the dataset `raw_analyst_ratings.csv`, which contained time-stamped financial headlines and related metadata (kag, 2025). I restricted the analysis window to **January 1, 2020 through May 31, 2020**.

To make the analysis efficient and reproducible, I sampled a fixed pool of **3,000** headlines using a fixed random seed (`NP_SEED=42`). I scored this pool once and reused the scored pool for event-based sampling.

3.2 Market data

I downloaded S&P 500 index data (`^GSPC`) using `yfinance` (Ranaroussi, 2025). I computed daily returns using the percentage change in closing prices and used returns (rather than raw prices) to reduce non-stationarity when computing correlations.

4 Methodology

4.1 Sentiment scoring

For each headline, FinBERT returned one of `{positive, negative, neutral}` with a confidence score. I mapped this output into a continuous signed sentiment score:

$$s(h) = \begin{cases} +\mathrm{conf}(h) & \text{if label is positive} \\ -\mathrm{conf}(h) & \text{if label is negative} \\ 0 & \text{if label is neutral} \end{cases} \quad (1)$$

This produced a score approximately in $[-1, 1]$ while preserving model confidence.

4.2 Events and balanced sampling

I analysed two event timestamps inside the study window:

- **WHO pandemic declaration** on March 11, 2020 (World Health Organization, 2020)
- **Federal Reserve announcement** on March 23, 2020 (Board of Governors of the Federal Reserve System, 2020)

For each event, I sampled **400** headlines from before the event and **400** headlines from after the event (balanced sampling). This design reduced sensitivity to imbalanced group sizes.

4.3 Hypothesis test and effect size

I compared sentiment distributions using a two-sided Mann–Whitney U test (Mann and Whitney, 1947). I also reported Cliff's delta (Cliff, 1993), a non-parametric effect size in $[-1, 1]$.

4.4 Market correlation

I aggregated the scored pool by date to compute daily mean sentiment and merged it with the daily return series. I then computed Pearson correlation between daily mean sentiment and daily returns:

$$r = \mathrm{corr}(\bar{s}_t, r_t) \tag{2}$$

I treated this analysis as exploratory; correlation alone did not imply causation or trading profitability.

5 Results

5.1 Event-based sentiment shifts

Table 1 reported the event comparison results. Both events showed statistically significant sentiment shifts.

5.2 Distribution visualisations

Figure 1 and Figure 2 visualised distribution shifts using violin plots exported by my script.

5.3 Market correlation

Daily mean sentiment was positively correlated with S&P 500 daily returns ($r = 0.3296$, $p = 0.0007$). Figure 3 plotted market close prices together with a 7-day rolling sentiment series for readability.

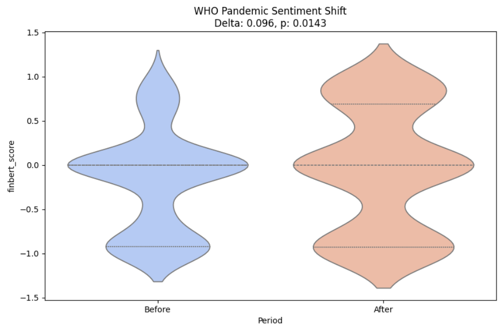

Figure 1: Sentiment distribution before vs. after the WHO declaration (balanced sampling).

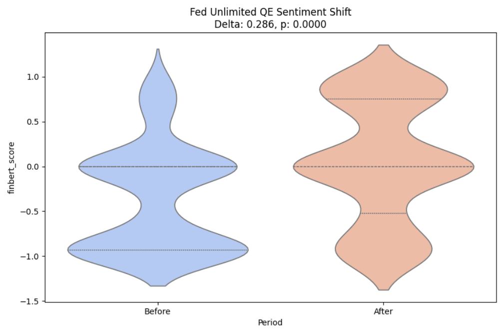

Figure 2: Sentiment distribution before vs. after the Federal Reserve announcement (balanced sampling).

6 Discussion

The results indicated that headline sentiment shifted around both events. The WHO declaration corresponded to a statistically significant but modest shift toward less negative sentiment. The Federal Reserve announcement was associated with a larger shift, consistent with the idea that strong policy signalling reduced perceived systemic risk during the same period.

The positive sentiment–returns correlation suggested that the extracted sentiment aligned with market movements during this specific time window. However, I did not interpret this as a predictive trading result because I did not run any forward-looking or out-of-sample forecast evaluation.

7 Limitations

This study had several limitations:

- I analysed titles only, which reduced context and increased ambiguity.
- The 3,000-headline pool was a sample and may not have fully represented the full news

Event	Mean Before	Mean After	p-value	Cliff's Δ
WHO declaration (Mar 11)	-0.1788	-0.0384	0.0143	0.0961
Fed announcement (Mar 23)	-0.3449	0.0465	4.93×10^{-13}	0.2863

Table 1: Balanced (400/400) comparisons of FinBERT sentiment before and after each event. p-values were computed using a two-sided Mann–Whitney U test.

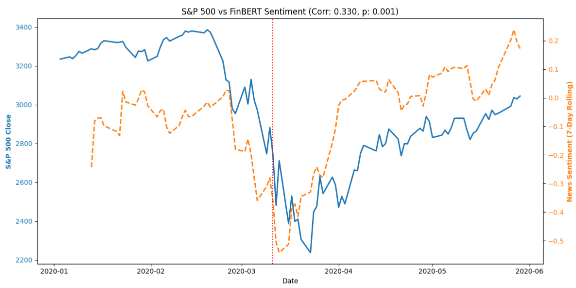

Figure 3: S&P 500 close prices and 7-day rolling FinBERT headline sentiment (Jan–May 2020).

stream.

- FinBERT may have model-specific biases; additional models and calibration could improve robustness.
- The correlation analysis was exploratory and did not establish causality.

8 Ethical Considerations

This project used publicly available text and market data. While the analysis could support market monitoring, it could also be misused to justify speculative claims. I reported results transparently and avoided presenting correlation as a guaranteed predictive signal.

9 Conclusion

I built a reproducible pipeline to analyse sentiment from financial news article titles and applied it to a 2020 market shock case study. The analysis found statistically significant sentiment shifts around two events and identified a positive association between daily sentiment and daily market returns during the same period. Future work could extend the approach to additional years and events, use full articles instead of titles, and evaluate predictive models with strict out-of-sample testing.

References

2025. Raw analyst ratings / stock news dataset (kaggle). https://www.kaggle.com/. Dataset source page varies by uploader; accessed 2025-12-29.

Board of Governors of the Federal Reserve System. 2020. Federal reserve announces extensive new measures to support the economy (23 march 2020). https://www.federalreserve.gov/. Accessed 2025-12-29.

Norman Cliff. 1993. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3):494–509.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

C.J. Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*.

Henry B. Mann and Donald R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60.

ProsusAI. 2025. Prosusai/finbert model card (hugging face). https://huggingface.co/ProsusAI/finbert. Accessed 2025-12-29.

Ran Ranaroussi. 2025. yfinance: Yahoo! finance market data downloader. https://github.com/ranaroussi/yfinance. Accessed 2025-12-29.

World Health Organization. 2020. Who director-general's opening remarks at the media briefing on covid-19 (11 march 2020). https://www.who.int/. Accessed 2025-12-29.

A Reproducibility Notes

I fixed the random seed (`NP_SEED=42`), cached FinBERT scores for a 3,000-headline pool, and saved event-specific sampled datasets, figures, and summary outputs to disk. The figures included in this report were exported by the same script used for scoring and evaluation.