Does News Sentiment Impact Stock Prices?

An Analysis on Nvidia-Related News and Stocks 2012-2022

Daniel Stevens
Dalarna University
Borlänge, Sweden
e-mail: h17daste@du.se

Omar Omran
Dalarna University
Borlänge, Sweden
e-mail: h24omaom@du.se

***Abstract*—We show that a selection of BERT-derived news headline sentiment for Nvidia has a positive correlation with the company's stock price, and that alongside the day's stock performance a simple logistic regression or decision tree can be used to predict a gain or loss in the next day's market open with a >50% accuracy. Further research is required to see whether the same holds true for other companies, for broader selections of news, and for longer-term prediction.**

Keywords- sentiment analysis, stock prediction, Nvidia

I. Introduction

This project investigates whether the sentiment of news headlines has an impact on stock price actions in the S&P 500, News performs a primary function in shaping investor perceptions, however, we intention to study if there's a measurable connection among headlines and short time stock performance.

Our core question is: Can the sentiment of daily news headlines provide an explanation or predict stock price changes? To solve this, we practice sentiment analysis techniques to a massive dataset of headlines and examine the results with matching stock information from 2010 to 2022.

The goal is to identify patterns that could link news sentiment with price movements and discover whether this information can offer important inquiries in financial analysis and by combining text mining and stock data, the project contributes to a better understanding of ways real world information affects market behavior.

II. Literature Review

Sentiment extracted from news headlines and social-media posts has been shown to affect equity-market moves. Heston and Sinha [1] demonstrate that firm-specific news polarity explains cross-sectional return variation across the S&P 500 even after controlling for past returns and market factors. Moving beyond dictionary methods, transformer models such as FinBERT deliver higher sentence-level accuracy, and Hu et al. [2] report economically significant information ratios when FinBERT scores drive trading rules on large-cap US stocks.

A. *Nvidia as a Case Study*

Event-study research finds sharp, statistically significant price reactions to sentiment-heavy news at the firm level. Nemes and Kiss [3] apply a BERT classifier to headlines and show that positive news produces +0.21 % abnormal returns for S&P 500 constituents on the announcement day, while negative news generates –0.18 %. For Nvidia specifically, Zheng and Feng [4] document cumulative abnormal returns of 20–30 % in the thirty trading days surrounding the ChatGPT-3.5 and 4.0 launches. Conversely, Reuters records a one-day 17 % loss after Chinese start-up DeepSeek unveiled a competing model in January 2025 [5] and a further decline tied to U.S. export-control headlines in April 2025 [6].

Academic work also evaluates real-time sentiment feeds: Bollen, Mao and Zeng [7] show that aggregate Twitter mood predicts daily DJIA direction with 86 % accuracy, while Liu and Srivastava [8] build a FinBERT dashboard that outperforms VADER on S&P 500 intraday returns. Together, these studies motivate our research question: does daily headline sentiment help explain or predict price movements across the S&P 500? We extend prior work by applying BERT to 4.5 million headlines for all index constituents (2010–2024) and testing both contemporaneous correlations and predictive power.

III. Method Description

A. *The Dataset*

Two datasets were used for this project. The first dataset [9] consisted of 4.5 million news headlines from "the top 10 news outlets by internet viewership", covering all sorts of unfiltered news from 2007-2022. To fit within our limited compute time, the dataset was filtered to only include the 455 headlines that contained the word "Nvidia" and further condensed into 232 daily mean sentiments after running a sentiment classifier. [10][11]

Note that in doing so, we likely cut many Nvidia-related headlines that may have used other keywords; however, we determined that collating a comprehensive list of keywords would have been enough for an entire study on its own, and more likely to include false positives than simply using the company name.

The second dataset [12] consisted of S&P 500 stock data from 2010-2024, which we filtered to only include the 2010-2022 NVDA prices. Columns describing the percentile change between days were added, making sure to cover both premarket, market, and post-market price changes.

These datasets were then merged with an inner join on date, which unfortunately resulted in the loss of weekend headlines as those are outside market days.

We choose four derived variables for our analysis: mean sentiment of that day, with -1 being negative, 0 being neutral, and +1 being positive; the percentile change from yesterday's close to today's open; the percentile change between the open and close of today's market; and the percentile change seen in the open of tomorrow's market.

We found that the correlations (Fig. 1) between the pre-market, daily market, and post-market were weak. News sentiment, however, had a stronger positive correlation with the postmarket prices and a slightly stronger positive correlation with the premarket prices. We assume that this may be due to the inclusion of the Daily Mail and BBC News in the dataset, as they have a four-hour head start on the UTC -4 New York Stock Exchange. As the granularity of our datasets is daily, some of the headlines are also created by the premarket changes. [13]

By plotting a linear regression over the premarket price (Fig. 2) we can see that it is positive within the default 95% confidence interval. We can also see that that the most extreme losses are only associated with negative mean sentiments, but again this may have the causation reversed due to those losses generating negative headlines later in the same calendar day.

B. Data Mining Method

In our naïve classification models, we assumed that the only known variables were today's premarket price change (percentile), today's price change from open to close (percentile), and the mean sentiment of the day's headlines (with negative being treated as -1, neutral 0, and positive 1). The class sought to be predicted is whether the stock value went up or down by tomorrow's open.

The data was randomly split into 70% training data and 30% testing data.

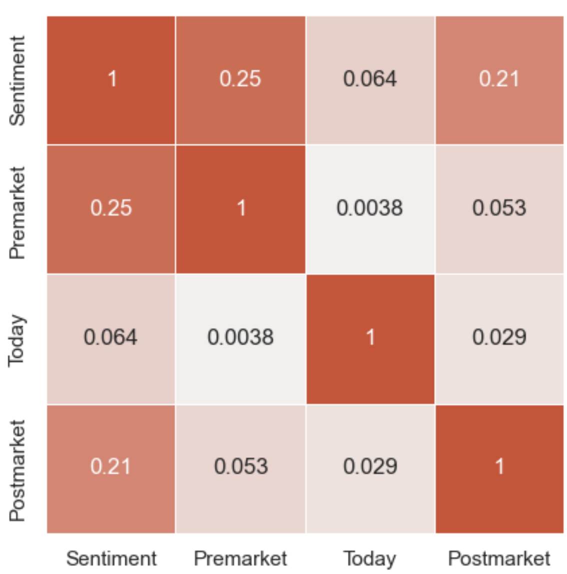

Figure 1 Correlation matrix for the combined dataset

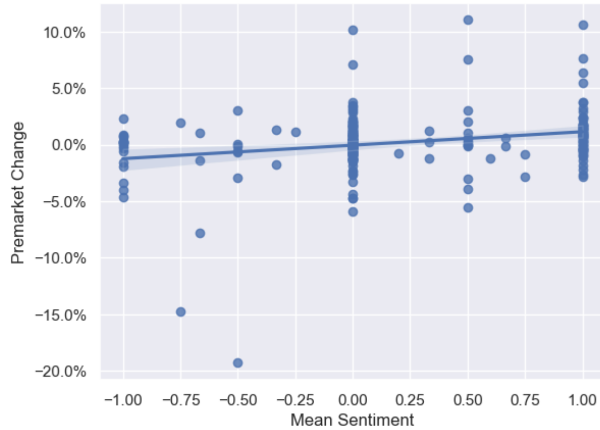

Figure 2 Linear regression of premarket prices and sentiment

For classification purposes, we ran our dataset through a logistic regression and a decision tree.

IV. Results and Analysis

The first classification algorithm we ran was a logistic regression, which achieved 62.86% accuracy in predictions and an F1 score of 0.76. However, when examining the confusion matrix (Fig. 3) it became obvious that it did so with a large bias towards predicting a gain. While this bullish strategy may be true for Nvidia's history as a rising stock, it seems a dangerous one as it guessed wrong on 88% of future losses.

We then ran a decision tree classifier on the same data (Fig. 4), which only achieved 57.14% accuracy and an F1 score of 0.66. The tree's confusion matrix (Fig. 5) shows that this reflects a 66% accuracy in predicting gains, which is smaller than the logistic regression, but a higher 46% accuracy in predicting losses.

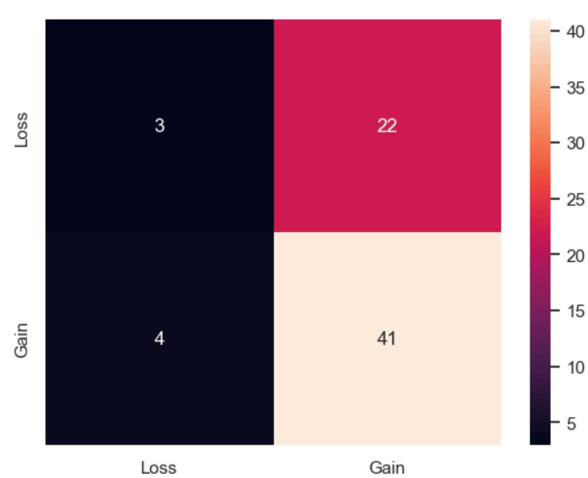

Figure 3 Confusion matrix for the logistic regression

V. CONCLUSION

In this project we explored whether the sentiment found in news headlines can affect stock prices, using Nvidia and the general S&P 500 as examples. Our results confirmed that headline sentiment has an obvious connection with stock price movements. What we noticed is that positive news mostly leads to price gains while negative news relates to declines. sentiment analysis, especially using advanced models like BERT, showed helpful in explaining daily stock price changes.

However, predicting future price movements using sentiment alone proved difficult. The models tested provided reasonable accuracy highlighting the complexity of stock markets.

For future research, we recommend that a better method for filtering headlines related to specific companies be found; that a more extensive study is done on perhaps the entirety of the S&P 500 as well as other markets; that higher-granularity time data be located to reduce questions of causality; and that time-lagged data be considered to investigate the possibility of long-term predictions.

REFERENCES

[1] S. B. Heston and N. Sinha, "News versus sentiment: Predicting stock returns from news stories," Rev. Financ. Stud., vol. 30, no. 10, pp. 3359–3411, 2017. [Online]. Available: https://doi.org/10.1093/rfs/hhx059

[2] Z. Hu, Y. Li, S. Cao, and W. Wang, "Deep-learning text mining for stock return prediction: FinBERT and beyond," IEEE Access, vol. 10, pp. 73728–73740, 2022. doi:10.1109/ACCESS.2022.3187791

[3] L. Nemes and A. Kiss, "Prediction of stock value changes using sentiment analysis of stock news headlines," J. Inf. Telecommun., vol. 5, no. 3, pp. 375–394, 2021. [Online]. Available: https://doi.org/10.1080/24751839.2021.1874252

[4] Y. Zheng and Y. Feng, "Impact of ChatGPT release on high-tech company: Evidence of NVIDIA's stock," Financial Engineering and Risk Management, vol. 7, no. 2, pp. 752–768, 2024. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4957042

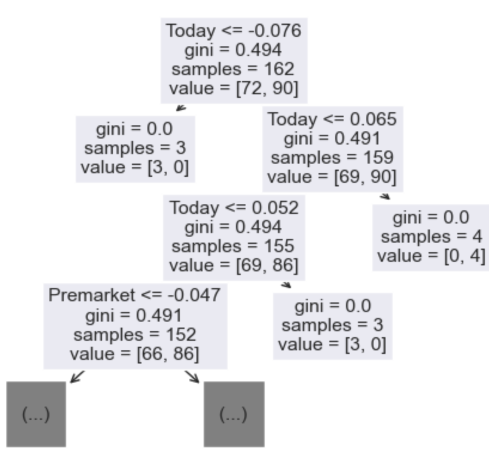

Figure 4 First three branches of the decision tree

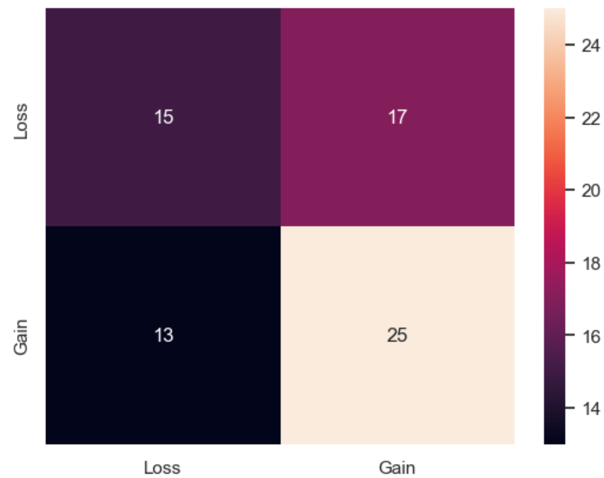

Figure 5 Confusion matrix for the decision tree

[5] Reuters, "DeepSeek sparks AI stock sell-off; Nvidia posts record market-cap loss," Jan. 28 2025. [Online]. Available: https://www.reuters.com/technology/chinas-deepseek-sets-off-ai-market-rout-2025-01-27

[6] Reuters, "Nvidia faces $5.5 billion charge as US restricts chip sales to China," Apr. 17 2025. [Online]. Available: https://www.reuters.com/technology/nvidia-faces-55-billion-charge-china-chip-curbs-2025-04-17

[7] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," J. Comput. Sci., vol. 2, no. 1, pp. 1–8, 2011. doi:10.1016/j.jocs.2010.12.001

[8] Q. Liu and A. Srivastava, "Benchmarking FinBERT dashboards for intraday stock prediction," Decis. Support Syst., vol. 167, 103991, 2023. doi:10.1016/j.dss.2022.103991

[9] Krishnayah, J. (2023). 4.5M headlines from 2007-2022 [10 largest sites]. *Kaggle.* https://www.kaggle.com/datasets/jordankrishnayah/45m-headlines-from-2007-2022-10-largest-sites

[10] Cardiff NLP (2022). Twitter-roBERTa-base-sentiment-latest. Huggingface. https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest

[11] Loureiro, D., Barbieri, F., Nevers, L. Anke, L.E., Camacho-Collados, J. (2022) TimeLMs: Diachronic Language Models from Twitter. DOI: https://doi.org/10.48550/arXiv.2202.03829

[12] Larxel. (2025) S&P 500 Stocks (daily updated). *Kaggle.* https://www.kaggle.com/datasets/andrewmvd/sp-500-stocks

[13] Haselton, T. (May 9 2017) Nvidia stock rockets up after earnings beat. CNBC. https://www.cnbc.com/2017/05/09/nvidia-fiscal-first-quarter-earnings-beat-expectations.html