Calculating Media’s Impact on Stock Price

Colby Wise & Flavien Moise

11/20/25

Motivation

- Financial markets are unpredictable
- Stock prices take into account many factors, from past performance to politics
- Media stories are often how these factors are reported

Question: how much impact do media stories have on stock price?

Background

- A share of **stock** is a portion of ownership in a company; owning shares of a stock means you own part of that business.
- The **price of a stock** changes based on how much buyers and sellers are willing to pay.
- Many things can affect how a stock's price moves, including:
  - Company news, financial results, or big announcements
  - Overall economic health, interest rates, inflation, and employment
  - Public opinion and how confident investors feel about the future
- **Sentiment analysis** is a computer technique that tries to measure how people feel, based on the words they use.
- **BERT** is a machine learning model that can understand context and subtle meanings, making it ideal for sentiment analysis

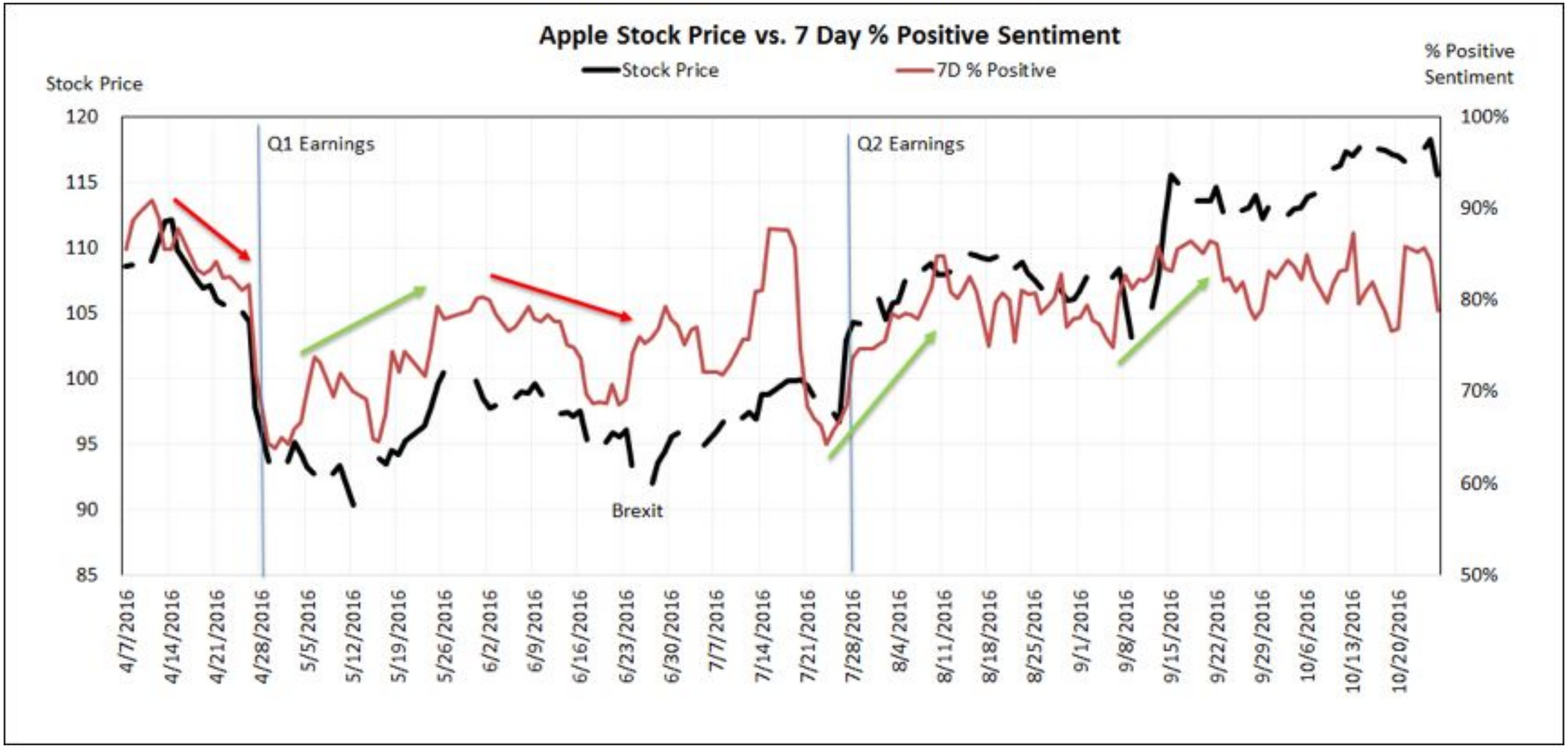
Related Work

- "BERTopic-Driven Stock Market Predictions: Unraveling Sentiment Insights" analyses comment sentiment and integrates it with deep learning models for stock price prediction.
  - https://arxiv.org/pdf/2404.02053
- "Market Trend Prediction using Sentiment Analysis: Lessons Learned and Paths Forward" examines how both sentiment attitude and emotion from financial news and tweets can, in certain cases, help predict market movements.
  - https://arxiv.org/pdf/1903.05440
- "A Sentiment Analysis Approach to the Prediction of Market Volatility" investigates correlations between sentiment from news headlines and tweets with next-day market volatility and returns, finding clear links in some cases.
  - https://arxiv.org/pdf/2012.05906

Claim / Target Task

- How much do media stories impact stock price?

An Intuitive Figure Showing WHY Claim



As can be seen in this graph, Apple's stock price is sometimes highly correlated with the sentiment analysis, and sometimes not. We proposed to do a multi-stock analysis to find a good overall correlation coefficient between sentiment and price.

Source: https://stocksnips.ai/learn/stock-news-sentiment-vs-stock-price/

Proposed Solution

1) Perform a sentiment analysis of media stories on specific stocks using a BERT model
2) Using statistics, determine the correlation between sentiment and stock price
3) Find the overall correlation coefficient between media sentiment and stock price

Implementation

Data: We will use a kaggle financial news dataset, which contains thousands of financial news articles from trusted outlets and their date of publication, and append financial stock info from finance for that day. (https://www.kaggle.com/datasets/rdolphin/financial-news-with-ticker-level-sentiment)

Sentiment Analysis: We plan on using finBERT, a BERT model on HuggingFace fine-tuned on financial data, to assign a sentiment analysis score for each article. (https://huggingface.co/ProsusAI/finbert)

Correlation: We can perform a linear regression to measure the relationship between sentiment and price change. An average & standard deviation of all stock correlations will be taken for our final answer.

Data Summary

- 13,386 headlines
- 2,833 missing/incomplete rows
- ~78% data is usable
- Data columns:
  - Headline
  - Date
  - Stock Ticker
  - Open Price
  - Close Price
  - Price Change
  - Price Change Percent

Extracted from raw Kaggle dataset

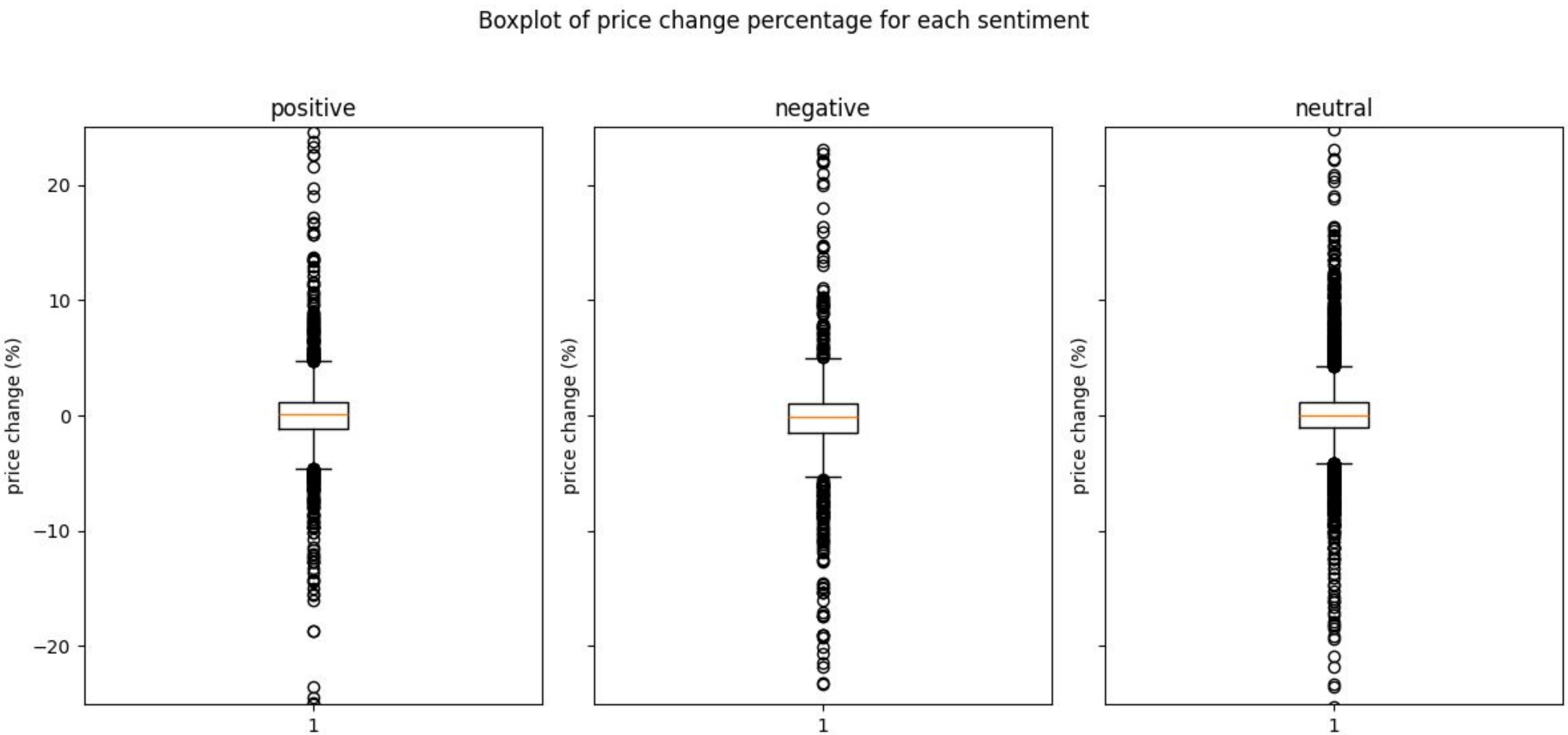
Added using yfinance API in python.

Experimental Results

	Correlation Coefficient	Mean % change in price	standard deviation of % change in price
BERT positive rating	0.0159	0.137	4.478
BERT neutral rating	-0.0016	0.104	3.568
BERT negative rating	-0.0158	-0.052	8.622

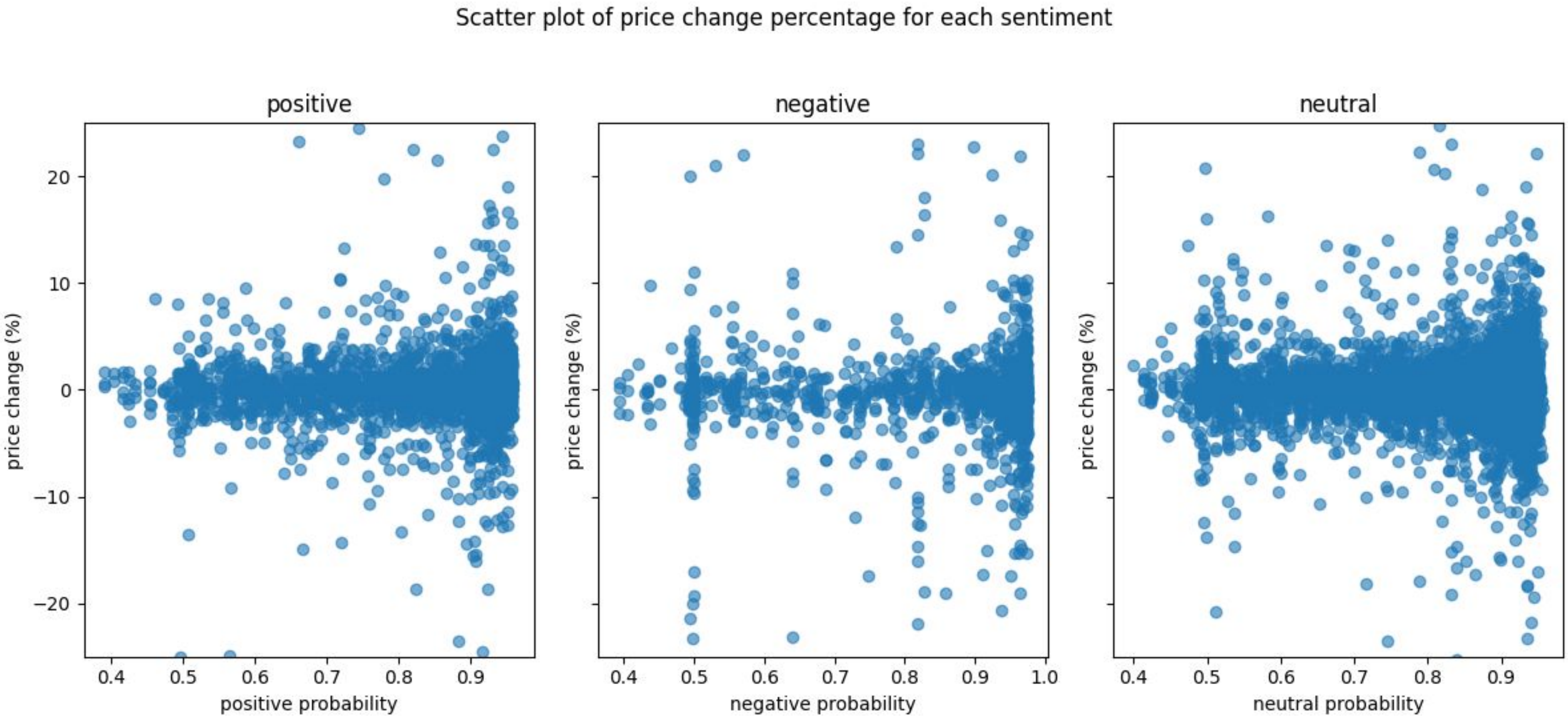
Key Takeaway: all the correlation coefficients are very low

Experimental Results



Key Takeaway: with all classes, all the data, except a few outliers, is near zero

Experimental Results



Key Takeaway: for most of the data, BERT was confident (high probability) in its classifications

Experimental Analysis

- Correlation coefficients are very low
  - All are under 0.02 - very weak!
- Mean for each class is near zero and standard deviation is large
  - Standard deviation for negative is unusually large
- Vast majority of headlines are classified as ‘neutral’
  - This could be a dataset peculiarity. Using another dataset with better headlines could be a solution.
  - This could also be a market issue: prices are too volatile, and media headlines are only one of several factors affecting price

Conclusion and Future Work

Conclusions:

- There is very little correlation between news coverage sentiment and stock price change by day
- Even a BERT model fine-tuned for finance does not discern good and bad news from neutral news very well

Possible Improvements:

- Include news article body or a summary in the input to BERT for inference
- Fine-tune a BERT model for each news source or journalist, to try and learn individual writing styles

Breakdown of Contributions

Flavien

- Performed research on topic, finding useful datasets, models, & previous work
- Did slides 2-9
- Wrote Python script for extracting data from the Kaggle dataset & enriching it with yfinance API

Colby

- Did slides 10-14
- Wrote Python Script for finding sentiment scores for each headline
- Wrote Python script for statistical analysis & creating the graphs

Due to different schedules & workloads, Flavien did more work towards the beginning of the project, & Colby did more at the end of it. Each person dedicated roughly the same amount of hours towards the project.

Code Demo Video:
https://drive.google.com/file/d/1_knZ4cJ5R_VtIu8JD36ALOirAj8IVv1H/view?usp=sharing

References

Deveikyte, J., Geman, H., Piccari, C., & Provetti, A. (2020). A sentiment analysis approach to the prediction of market volatility. *arXiv:2012.05906* [q-fin.ST]. https://doi.org/10.48550/arXiv.2012.05906

Mudinas, A., Zhang, D., & Levene, M. (2019). Market trend prediction using sentiment analysis: Lessons learned and paths forward. *arXiv:1903.05440* [cs.CL].

Zhu, E., et al. (2024). BERTopic-driven stock market predictions: Unraveling sentiment insights. *arXiv:2404.02053* [cs.CL]. https://doi.org/10.48550/arXiv.2404.02053