

# Sentiment Analysis Usage Within Stock Price Predictions

Ethan Chan, Wing Chan

*Received October 01, 2024*

*Accepted January 31, 2025*

*Electronic access February 15, 2025*

The growth of the United States economy in recent years has led to more investments in the stock market. However, the unpredictable nature of the stock market makes it difficult for new investors to forecast stock price trends. This study was conducted to determine if relevant news articles have an effect on stock prices, and if so, can sentiments from these articles serve as an input to investment decisions. A sentiment analysis model using the TextBlob library was created to collect the sentiments from the top 25 stocks with the highest market cap as of July 7, 2024. The 50 most recent articles in The New York Times newspaper of those companies were analyzed through the model, and the sentiment values of the articles and stock prices were found on matching date ranges. The data was then analyzed using Pearson's correlation and in multiple different groups to find possible trends over various categories; these included analyzing based on the timeframe, industry, growth percentage, and volatility. The results revealed a lack of significant correlation between the sentiment and stock prices, with correlation coefficients ranging from, at most, -0.5 to 0.5 and averaging near zero. Sentiment analysis, while a useful tool, is not capable of predicting stock prices alone. However, it can be used as a basic input for investment decisions, and may be further strengthened when used in collaboration with other AI models to yield more accurate results.

## Introduction

Understanding the factors influencing stock market fluctuations is crucial for developing effective investment strategies and mitigating financial risks. As the United States economy grows, an influx of new investors is entering the market, increasing the demand for more accurate predictions of stock movements. The complexity and rapid pace of the stock market have driven interest in AI-focused prediction models, which offer relative ease of use and fair accuracy. Common AI models include neural networks, natural language processing, pattern recognition models, and various other types. These models are the focus of current studies to optimize accuracy and reliability, offering promising predictions for stocks. However, the increase of sentiments expressed in news and other media outlets adds a possibility of influencing investor decisions. This highlights the importance of understanding how news can impact stock prices, especially as many investment decisions are influenced by public perception and information. This study explores the possible usage of sentiment analysis on news articles to find a cohesive pattern or trend.

A popular method of stock prediction using AI involves using convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Convolutional neural networks are a type of artificial intelligence that is well-suited for processing data in a grid-like structure, most commonly being images. Similarly, recurrent neural networks are another type of artificial intelligence, yet they process sequential data, being better acclimated to processing text, language, and time-based data.

These models analyze historical data to identify recognizable patterns within each stock's performance. CNNs and RNNs are compatible with time series data and can model complex, non-linear relationships, making them appear promising. Their ability to handle large datasets and uncover intricate patterns makes them well-suited for this task. However, these models are prone to overfitting, leading to increasingly inaccurate predictions over time. Furthermore, historical performance is not necessarily indicative of future performance. They also struggle to forecast anomalies and often fail to account for external factors affecting the economy and specific companies<sup>1</sup>. More complex and intricate neural networks or other AI models may be able to mitigate these problems.

Studies utilizing sentiment analysis, a method of using AI to determine whether written content is positive, negative, or neutral, to track and forecast stock price shifts have shown some promising results. Social media platforms, particularly Twitter, have been employed to study the efficacy of sentiment analysis in predicting stock prices. In previous studies, using social media has provided some correlation between sentiment and stock prices, with fairly accurate predictions when compared to actual price fluctuations<sup>2</sup>. However, this method may still face limitations in regards to social media harboring overly biased user opinions and rampant misinformation. Therefore, using credible news outlets can be a more accurate and dependable method for estimating future stock prices. This would capture investor behavior and market reactions to news, making it a useful tool for predicting stock prices<sup>3</sup>. Unlike social media sentiment, which can be biased and influenced

---

by misinformation, news articles are typically authored by professional journalists aiming to maintain objectivity. This makes them a potentially more reliable source for analyzing how news sentiment correlates with stock price movements. Generally, positive news can lead to an increase in stock prices, while negative news can cause a decline. By quantifying sentiment from news articles and correlating it with stock prices, the impact on market behavior can be found. This study aims to shift the focus to news articles published by credible outlets, which are typically authored with objectivity in mind. By analyzing the sentiment from reputable sources, this research seeks to address the challenges posed by the noise and unreliability of social media data. In addition, to find deeper patterns, this study will perform a sub-group analysis by analyzing data based on timeframes, industries, growth rates, and volatility. This approach furthers a new avenue for which a more comprehensive understanding of the potentials and limitations of sentiment analysis can be achieved.

As previously discussed, AI models have their limitations and drawbacks, and sentiment analysis is not an exception. While one benefit of using sentiment analysis on news outlets is its relative neutrality compared to social media platforms, this advantage has its caveats. For instance, the editor of a major news outlet or the publication as a whole could exhibit bias towards certain companies, skewing results and introducing favoritism. Averaging articles from multiple media sources could solve this issue, although this study does not cover that approach. Furthermore, macroeconomic events could dilute correlations with stock prices. Economic forces, such as recessions or geopolitical conflicts, may overshadow the effect of news sentiment on stock prices. Industry-specific developments or regulations may also have a similar effect. Despite these limitations, the findings of this study provide valuable insights into the relationship between news sentiment and stock price movements, offering a more controlled perspective compared to analyses based on social media data.

## Methods

This study adopts an empirical approach, finding and analyzing data to uncover patterns or trends. For this paper, the 50 most recent articles of the top 25 stocks based on the market cap, the total dollar value of a company's outstanding shares of stock, were collected on July 7, 2024. Only the 50 most recent articles were used due to a limitation involving retrieving articles using the New York Times API. Due to the amount of news coverage differing for each company, the time frame covered by the articles is different by company - and this is taken into account in the subgroup analysis. A broader market indicator, the SPDR S&P 500 (SPY), will be used as a baseline to compare company specific results. As opposed to using smaller companies, which could be biased due to outliers, choosing the top 25 stocks by

market cap would be beneficial due to how these companies are generally large and influential, often setting trends for the entire market. In addition, these companies have more news coverage, allowing for more data to be collected, increasing the reliability of the analysis. Although influential, the limited company does impose some the level of detail and comprehensiveness the study will explore. Having July 2024 as the endpoint for data collection ensures that the analysis reflects the most up-to-date trends and market conditions at the time of the analysis. While including more data from earlier periods (beyond the 50 articles) could provide insights into long-term trends, the long term historical data are assumed to have been taken into account in the historical stock prices. Future research could explore a broader historical dataset to determine whether the trends observed in recent data persist over longer periods.

Sentiment values for articles were found using a sentiment analysis model<sup>4</sup>, constructed primarily using the TextBlob library in Python. TextBlob was selected due to its accessibility and fair accuracy. While a well-known limitation with this model is its inability to account for sarcasm and other language dependent contexts, due to how it analyzes each word instead of entire phrases or sentences, the articles analyzed by the model did not face such limitations as they came from The New York Times, a reputable news source, which tends to be more straightforward in their language and less likely to use sarcasm. It is important to note that relying solely on a singular article could introduce a potential bias. Different news outlets, such as Reuters or Bloomberg, may have varying perspectives, which could influence the articles they cover and their tone surrounding the topic. However, for this study, The New York Times was chosen as the primary source due to its extensive coverage of financial news, high journalistic integrity, and its generally wide influence and popularity among citizens. Furthermore, the use of a single source allows the analysis to be more controlled and focused. While this model was not explicitly validated on manually labeled datasets, TextBlob has demonstrated its reliability, as shown in both academic and personal projects. Its lexicon-based approach, basing its evaluation of sentiment on pre-trained datasets, allows it to effectively analyze long text structures, including the articles analyzed in this study. This consistent lexicon-based framework highlights its reliability, as all projects utilizing TextBlob operate under the same foundation. However, to ensure consistency, the model was used on the articles multiple times to check for consistent results.

After data collection, sentiment values were multiplied by 100 to make the values less abstract, placing them between -100 and 100. This does not affect the analysis; it simply rescales the values without changing their relative magnitudes, making them appear simpler. The sentiment values of each article was matched with their historical stock price data. The dates of the articles were listed alongside their corresponding

sentiment values, both being reported by the model. Stock prices were obtained from NASDAQ and placed in another column. If multiple articles were released on the same date, their sentiment values were averaged. This alignment allowed for a straightforward process to compare both stock prices and sentiment values based on date. Pearson's correlation, a statistical method that quantifies how closely two variables are related, was used to measure the strength and direction of the relationship between sentiment values and stock prices. Pearson's correlation was chosen to analyze the relationship between stock prices and sentiment due to its ability to effectively measure linear associations. While other methods, specifically Spearman's correlation, might capture rank-based relationships well, this study focuses on both the direction and magnitude of the changes between sentiment and stock price, which Pearson's method addresses directly and Spearman's does not. Although more complex methods, such as regression models, could capture non-linear factors, they exceed the scope of this study and introduce additional complexity, though future studies may benefit from analyzing this aspect.

A correlation coefficient closer to 1 indicates a strong positive correlation, while a coefficient closer to -1 indicates a strong negative correlation. A value approaching 0 has little to no correlation. This approach would effectively estimate the correlation between stock prices and sentiment values, as rising sentiment would suggest a proportional increase in stock price. Any correlation near 0 would be considered to not correlate. In cases where the dataset contained an odd number of datasets, the middle date was included in both the first and second halves to ensure consistency and accuracy in the analysis. Correlation values would be the metric statistically analyzed to determine whether a stock has a recognizable pattern or not. The correlation values were organized into tables, which showed the data based on how they were going to be analyzed, being based on time periods, industry, percent growth, or volatility. For the time periods section, companies are categorized by the length of their analysis period, determined by the time horizon over which their articles were collected. For instance, AbbVie is classified as "very long" since its 50 articles span approximately five years, while Amazon is categorized as "short" because its articles spanned over a two-week period. For the volatility category, the beta values of each stock, a measure of how volatile they are, were found using Yahoo Finance and categorized by their magnitude from the value of 1. The beta values found for this paper were 5 year monthly beta values, meaning they were calculated using data collected monthly over the span of 5 years. By analyzing stocks within specific industries, sector-specific news, events, and performance can be controlled. Furthermore, categorizing stocks by growth percentage can help to control for other factors that may influence trends, allowing for a more isolated analysis of sentiment's impact on different growth groups. While these methods of analysis may have their own

limitations, they provide a valuable framework for addressing the nuanced relationship between news sentiment and stock prices.

Table 1. Listing of companies analyzed time periods, industry, percent growth, and volatility categories.

Listing of Companies (Table 1)				
Company	Length	Sector	Growth	Volatility (Beta)
AbbVie	Very Long	Pharmaceuticals and Healthcare	Very High	Low
Amazon	Short	Manufacturing and Consumer Goods	Low	High
Apple	Short	Technology	Low	High
Bank of America	Short	Financial	Low	High
Berkshire Hathaway	Very Long	Pharmaceuticals and Healthcare	Moderate	Low
Broadcom Inc.	Very Long	Technology	Very High	High
Costco Wholesale	Very Long	Manufacturing and Consumer Goods	Very High	Low
Eli Lilly and Co.	Very Long	Pharmaceuticals and Healthcare	Very High	Low
Exxon Mobil	Long	Energy	Moderate	Low
Google	Short	Technology	Low	High
Home Depot	Long	Manufacturing and Consumer Goods	Negative	High
Johnson & Johnson	Short	Pharmaceuticals and Healthcare	Negative	Low
JPMorgan Chase	Long	Financial	Low	High
Mastercard	Very Long	Financial	Moderate	High
Merck & Co.	Very Long	Pharmaceuticals and Healthcare	High	Low
Meta	Medium	Technology	Low	High
Microsoft	Medium	Technology	Low	Low
Netflix	Short	Technology	Low	High
Nvidia	Long	Technology	High	High
Oracle	Long	Technology	Moderate	High
Procter and Gamble	Very Long	Manufacturing and Consumer Goods	Moderate	Low
Tesla	Medium	Energy	High	High
UnitedHealth Group	Very Long	Pharmaceuticals and Healthcare	Moderate	Low
Visa	Medium	Financial	Negative	Low
Walmart	Long	Manufacturing and Consumer Goods	Low	Low

The timeframes chosen for this study (short-term being 2 weeks and the medium-term being 1 month, etc.) may not precisely align with traditional trading or investment horizons, such as intraday, weekly, or quarterly intervals. However, they were selected based on the nature of the available data and how stock performance is commonly analyzed on financial platforms. For instance, the medium-term timeframe reflects the standard 1-month view offered by many platforms, while the long-term and very long-term categories correspond to commonly used year-to-date, at least at the time of data collection, or 5-year views. Additionally, the timeframes were applied to the data from the New York Times API, which provided sentiment snapshots rather than continuous data, allowing broader trends to be easily analyzed over these time periods. The collected

sentiment and stock price data were analyzed to identify potential correlations. However, there are some factors, such as macroeconomic events, that can introduce limitation to this methodology.

It is also noted that there can be limitations to the analysis due to macro economic factors. During the period analyzed, there were some economic and geopolitical changes. For example, the cooling inflation rates from 6.8% in 2023 to 5.9% in 2024<sup>5</sup> may have created a more optimistic economic outlook. The general increase of the GDP of many of the world’s major trading nations, such as China and the United States, could have led to greater sentiment, as multinational companies would have better performance, therefore increasing the sentiment of articles reporting on those companies. On the other hand, political tensions, such as the conflict in Ukraine, likely decreased the sentiment for some parts of the market. Furthermore, the energy sector can experience fluctuation due to the emergence of renewable energy sources and technologies. The recent advancements and implementation of AI within many of the companies analyzed, not limited to technology-focused businesses due to its wide-ranging applications, may also affect sentiments. These broader events during the analysis period may have influenced the sentiments of the articles, which should be taken into context.

## Results

This study examined sentiment analysis of leading companies across multiple categories to find potential trends within these groupings. The data found was categorized by timeframes, industries, percent growth, and volatility to control for factors other than sentiment. Additionally, a baseline comparison was conducted using the SPDR S&P 500 (SPY) ETF, which measures the performance of the S&P 500 index<sup>6</sup>. The baseline serves as a reference point, to which the data can be compared to and find a distinction between specific sentiment correlations and general market behavior. The study found that there are weak correlations in some specific instances, but overall correlations were close to 0. In the data as a whole, the correlations ranged from -0.49 to 0.45, with a range of p-values from 0.022 to 0.994. However, there were three companies with p-value < .05, which is significant, being Berkshire Hathaway, Tesla, and Walmart, with p-values of 0.004, 0.022, and 0.027 respectively.

### Analysis by Time

The data in the table was categorized by time frames, which were the short term (2 weeks), medium term (1 month), long term (1 year), and very long term (6 years). The data from this short-term analysis reveals overall correlations ranging from -0.4 to 0.45. In the short term, sentiment analysis shows balanced results between being negative, positive, and having no correlation.

Time Intervals	Number of Companies	Min Correlation	Max Correlation	Average Correlation	P Value Range
Short Term	6	-0.4	0.45	0.02	0.123 - 0.929
Medium Term	4	-0.49	-0.03	-0.27	0.022 - 0.458
Long Term	6	-0.35	0.07	-0.03	0.027 - 0.65
Very Long Term	9	-0.39	0.36	-0.01	0.004 - 0.994

Summarized results of Table 2-5. Compares correlation data with the time span measured the companies. Larger, detailed company -by-company correlations can be found in Table 2<sup>7</sup>, Table 3<sup>8</sup>, Table 4<sup>9</sup>, and Table 5<sup>10</sup>.

The medium-term displays correlations ranging from -0.49 to -0.03, with the long-term showing a similar range, only shifting more positively. The medium-term shows more negative overall correlations when compared to the rest of the sections in this dataset. The long-term data shows very slightly positive near-zero correlations in most of the overall values, with the only exception being a moderately negative value. The very long-term group contains overall correlations ranging from -0.39 to 0.36. Once again, this dataset shows mixed values, similar to the short-term section.

### Analysis by Industry

Another approach to analyzing the data is to explore possible patterns within specific industries to determine whether any type of company may tend to have a high or low correlation. The major results are shown below.

Industry	Number of Companies	Min Correlation	Max Correlation	Average Correlation	P Value Range
Technology and Online	8	-0.18	0.45	0.07	0.184 - 0.929
Financial	4	-0.37	0.04	-0.17	0.055 - 0.602
Manufacturing and Consumer Goods	5	-0.4	0.37	-0.09	0.027 - 0.994
Pharmaceuticals	6	-0.5	0.23	-0.04	0.004 - 0.831
Energy	2	-0.49	-0.01	-0.25	0.022 - 0.466

Summarized Table 6. Shows the correlation between company sentiment and stock price for different industries. More information can be found in Table 6<sup>11</sup>. This only includes more major information about each company, as all data from each company can be found in Tables 2-5.

The technology industry had overall correlations ranging from -0.18 to 0.45, which is one of the more generally positive ranges over this dataset. It had mostly no correlation, with the few exceptions offsetting each other, producing such a near-zero average correlation. The entire financial industry



had much more negative correlations overall, ranging from -0.37 to 0.04, contributing to its fairly negative average correlation. Manufacturing and consumer goods companies also have a somewhat balanced correlation range, while its average is slightly negative. The pharmaceutical and healthcare industry has a more negatively skewed correlation range, extending between -0.5 and 0.23; however, due to its balanced correlations, it has a relatively zero correlation, resembling previous industries. The energy sector also has a very negative correlation range compared to other industries here, with the lowest average correlation.

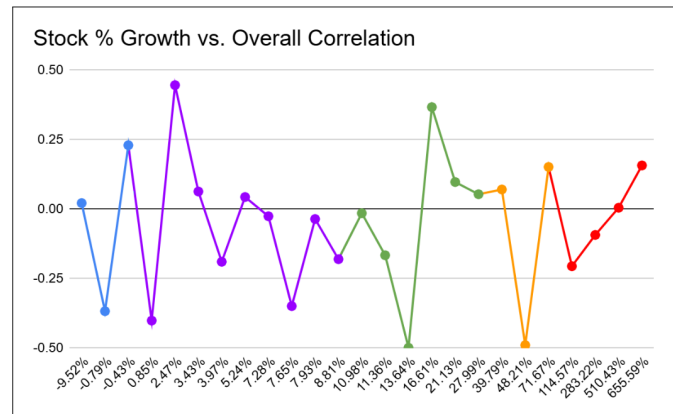
### Analysis by Percent Growth

Finally, the companies can be organized in a way based on the percentage by which they grew throughout the data collection period.

Correlation and Price Percentage Increase by Sentiment Values						
Company Sentiment	Number of Companies	Min Correlation	Max Correlation	Average Correlation	Average % Increase	P Value Range
Negative	3	-0.36	0.23	-0.04	-3.58%	0.054 - 0.692
Low	9	-0.4	0.45	-0.07	5.29%	0.027 - 0.929
Moderate	6	-0.5	0.37	-0.03	16.95%	0.004 - 0.994
High	3	-0.49	0.15	-0.09	53.22%	0.022 - 0.831
Very High	4	-0.21	0.16	-0.03	390.95%	0.113 - 0.86

Summarized Table 7. Shows the correlation between company sentiment and stock price percentage growth. Further details about this data can be found in Table 7<sup>12</sup>. Full details about each company's sentiment and stock price based on each date analyzed can be found in Tables 2-5.

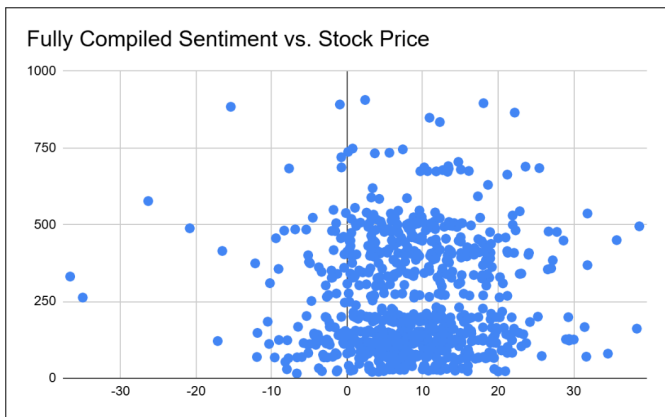
The negative growth category has correlations ranging from -0.36 to 0.23. Its mixed correlation also contributes to the near-zero correlation. The low-growth section has a mostly balanced correlation range, and the overall correlations, though generally lower, remain close to zero. The moderate growth category is similar to the low growth section, where the lower bounds of the ranges are more negative, and overall correlations are fairly balanced. The high growth data has a negatively leaning range of -0.49 to 0.15, with a slightly negative average. The very high growth category has a moderately balanced correlation range, from -0.21 to 0.16, with a very near-zero overall average correlation. The very high growth group is also shown to increase in correlation as percentage growth also increases, as shown in Figure 1. However, the figure does not show a similar relation anywhere else, oscillating between negative and positive values. This confirms how higher stock percentage growth is not typically indicative of a higher correlation.



**Fig. 1** Line graph comparing stock percentage growth on the x-axis to overall correlations on the y-axis. Multiple points leading in one direction indicate more correlation, whether positive or negative, as the percentage increases. The blue points represent negative growth, the purple points represent low growth, the green points represent moderate growth, the orange points represent high growth, and the red points represent very high growth. The colored lines in between each of the points are not indicative of anything.

Correlation by Beta						
	Number of Companies	Min Correlation	Max Correlation	Average Correlation	Average Beta Value	P Value Range
Low	12	-0.39	0.37	-0.05	0.66	0.004 - 0.994
High	13	-0.49	0.45	-0.05	1.28	0.022 - 0.929

Summarized Table 8. Shows the correlation between company sentiment and volatility. Further details about this data can be found in Table 8<sup>13</sup>. Full details about each company's sentiment and stock price based on each date analyzed can be found in Tables 2-5.



**Fig. 2** The figure shows the components that make up the correlations for the stocks. Each point represents a sentiment value (x-axis) and its stock price for the day the article was published (y-axis). This scatterplot shows the compiled plot of all of the articles analyzed across all of the companies.

### Analysis by Volatility

The beta category shows the beta value of each stock analyzed. The high volatility category shows that it had both the lowest and highest correlations. Therefore, due to such polar correlation values, the correlation averages to about 0. Similar happens in the low volatility category, where there are equal magnitudes of positive and negative values, creating another average correlation of close to 0. The average beta value of the low category is quite far from 1, while the high category is only moderately away from 1.

The scatterplot supports that the overall results revealed weak and inconsistent correlations between sentiment values and stock price movements across various timeframes, industries, and growth categories. Despite some individual examples showing slight trends, the majority of the data exhibited correlations close to zero, suggesting the limited predictive capability of sentiment analysis in isolation. These findings are explored in greater detail in the following discussion.

### Discussion

The results of this study highlight the complexities of using sentiment analysis to predict stock price movements. While some weak correlations were observed, such as within specific examples in the timeframes and industries subgroup analysis, the data revealed no consistent patterns. This inconsistency suggests that sentiment analysis may be able to capture some of the factors influencing stock prices but not all. The scatterplot further supports this, where it was revealed that there was minimal overall growth, both for individual companies and as a whole. This absence of a defined trend further emphasizes how there was only little correlation between stock prices and sentiment.

Overall, the data suggests that sentiment analysis found limited patterns of correlation with stock price movements across different time periods. In the short term, sentiment analysis shows mixed results, with correlations ranging from negative to positive, indicating no clear trend. This inconsistency can be caused by heightened market volatility, where impulsive reactions to news or rumors create rapid fluctuations that are reflected in stock prices. The lack of uniform correlation can also be attributed to the multitude of factors influencing stock prices that sentiment analysis does not account for, such as day-to-day fluctuations and differing investor sentiment, which may not align directly with news sentiment. As time progresses, the relationship between sentiment and stock prices becomes even more nuanced and less predictable. For example, the medium term shows more negative correlations, while most stocks in the long term show no correlation. The very long term shows balanced results again, similar to the short term. Over time, long-term data analysis suggests that larger economic forces, such as market cycles and industry developments, dilute the influence of short-term factors. As data shifts from being in the short term to the long term, the stronger and more dominant features of the market diminish the impact of short-term developments. These external economic and market conditions makes the sentiment analysis in the long term less useful.

Across industries, little consistent patterns can be found between sentiment analysis and stock performance, except for some specific company examples. In the technology sector, correlations hovered around zero, as its diverse range of companies and services results in offsetting sentiments. However, Netflix stood out with a relatively high correlation of 0.446 (p-value = 0.929), suggesting that news sentiment may have a stronger influence on its stock price. This could be because of Netflix's position as a consumer-facing company heavily reliant on public perception about its content releases. Conversely, companies like Meta and Microsoft exhibited correlations close to zero, indicating that their stock prices may be driven by other factors, such as product advancement, rather than short-term sentiment. The general lack of clear correlation within this industry could also be due to the ongoing debates about the ethical and practical aspects of technology, leading to more varied public opinions. The financial sector, however, showed a general leaning toward negative correlations. This trend may reflect economic concerns such as inflation, rising interest rates, and general uncertainty, which heavily impact financial institutions. For instance, Visa displayed a relatively high negative correlation of -0.368 (p-value = 0.055), highlighting its sensitivity to economic indicators like consumer spending and credit availability, which may be negatively represented within news sentiment. This cause does not significantly affect stock prices, although it leads to a decline in sentiment, causing the negative correlation. Overall, this industry's widespread usage offers stability and a generally slow

---

rise in stock price, which, when paired with negative sentiment, could lead to a negative correlation between sentiment and stock prices.

Manufacturing and consumer goods companies exhibit balanced correlations, with sentiment impact most likely being diluted by their diverse product ranges. Procter & Gamble demonstrated a moderate positive correlation of 0.367, while Amazon and Walmart exhibited negative correlations, being -0.402 and -0.349, respectively. This high disparity within a single industry may indicate that their performance might be more strongly driven by market trends and operational factors rather than short-term news sentiment. Similarly, the pharmaceutical and healthcare industry exhibit balanced correlations, where sentiment seems less impactful compared to company-specific factors and industry-wide changes, such as regulatory reforms or medical breakthroughs. Eli Lilly and Co, for example, has had many of their drugs recently approved by the FDA, though that positive news is overshadowed by other recent and uncertain healthcare reforms. Regarding correlations, however, Johnson & Johnson had the highest correlation within this sector of 0.230, which may reflect its sensitivity to news regarding healthcare developments or consumer trust. On the other hand, AbbVie's negative correlation of -0.206 suggests that its stock performance may be more influenced by factors, such as its company's drug approvals and regulatory developments, rather than short-term sentiment shifts in the news. The energy sector displays more negative correlations, but with such a small sample size, the data might not accurately reflect broader trends in the industry. Tesla's strong negative correlation of -0.490 (p-value = 0.022) simply indicates that its sentiment does not follow its progressive rise in stock price, possibly due to controversy surrounding the company. Conversely, ExxonMobil's near-zero correlation of -0.015 suggests that its stable stock price is not affected by its fluctuating sentiment. Instead, it is more likely to be driven by factors like fluctuating oil prices and the increasing use of renewable energies, which have a greater impact on ExxonMobil's stock price than the sentiment expressed in news coverage. A more diverse selection of companies can be useful to draw reliable conclusions within that industry. Ultimately, stock performance seems to be more influenced by individual company events rather than industry-wide sentiment trends.

The stocks analyzed based on their growth percentages were sorted into several growth groups: negative, low, medium, high, and very high. The theory behind this way of categorizing is that higher growth percentages might indicate stronger correlations, though the data does not prove this theory. In the negative growth category, correlations are mixed between being negative, positive, and having no correlation, meaning there is no observable pattern here. Low-growth stocks mostly exhibit slightly negative near-zero correlations, though there is no dominant pattern that dictates their low growth. The moderate growth category resembles the low-growth group, with a mix

of near-zero and slightly negative correlations, again showing no clear trend. High-growth stocks lean towards negative correlations, despite their comparatively high growth. The very high growth category also presents balanced correlations, with an overall average close to zero. There is some indication that correlations increase as growth percentages rise, but more data would be required to verify this point.

When analyzing by each company's beta, the groups were separated into high and low categories. The high beta group contains companies whose 5-year beta was greater than one, while low meant that the companies' beta values were below one. In contrast, the beta of the benchmark S&P 500 (SPY) was one. Within the high volatility group, correlations between stock prices and sentiment have much greater magnitudes, while there was still a roughly equal mix between negative and positive correlations. Higher beta stocks are more sensitive to market movements and external factors, including public sentiment expressed within news. These companies often react fairly dramatically to both positive and negative developments, reflecting their riskier nature. This sensitivity magnifies the impact of sentiment on stock prices, as investors are quicker to react to market news. This drives correlation greatly positively or negatively, depending on the news sentiment expressed. As the beta values of each company approach zero, the correlations tend to center around zero as well. However, aside from Visa's relatively high correlation of 0.368 while still having a beta value of 0.95, lower volatility stocks showed a general increase in magnitudes as the beta values grew further from 1, again having mixed negative and positive correlations. Such high correlations in lower volatility comes unexpected at first, especially since they are expected to be more stable and therefore have lower correlations. Due to the fact that it is generally believed that these stocks would be more stable, this makes those companies more reactive, and potentially shocked, by an unexpected shift in sentiment. This would then allow the stock prices to be more affected by news sentiment, producing the high correlations seen. Not only do stocks with higher volatility correlate better with their sentiment and price, but stable stocks do as well. Finally, stocks with beta values surrounding the S&P 500's beta value have lower correlations expected in stable stocks.

Across all categories, the lack of a consistent trend suggests that stock price growth alone is not a reliable predictor of correlation trends. The more a stock grows does not necessarily indicate it has a higher correlation. Other economic factors and company-specific influences may have a stronger impact on stock performance than sentiment analysis alone can capture. Additionally, the small sample size and longer time frames used in the analysis limit the strength of these findings, needing further research with a larger dataset. The comparison of stock price growth with the correlation between stock prices and sentiment values may also suggest that sentiment has more of an effect on correlation values. However, due to fluctuating

---

correlation values, refining sentiment analysis with more data or improved models is essential for better accuracy. Continued innovation using sentiment analysis as a supporting tool may be more beneficial instead of using sentiment analysis alone.

In comparison to the individual companies, the baseline SPDR S&P 500 (SPY) ETF, representing broader market performance, exhibited an overall correlation of -0.121, with a p value of 0.255. For a broad market indicator like the S&P 500, there are more diverse, macro reasons influencing the stock price. While the sentiment analysis may not yield statistically significant results, the SPY analysis does serve as a good benchmark and context to which the individual companies' results can be used to compare.

## Conclusion

At first glance, sentiment analysis holds promise in associating stock prices with news media to uncover patterns that might be useful for predicting future stock movements. The objectivity of sentiment analysis, paired with its news coverage, offers some advantages over other models. However, this approach has clear limitations. One of the limitation lies in the relatively small dataset and the limited range of news sources considered in this study. Expanding the analysis to include a larger number of companies would likely yield more comprehensive results. While sentiment analysis provides useful insights, it falls short of fully capturing the factors that influence stock prices.

Nonetheless, the findings from this study offer promising and practical applications. Investors and analysts could use sentiment analysis as a supplementary tool to support other strategies rather than as the primary mechanism. For example, investors might rely on sentiment analysis to measure market reactions to significant news events, such as earnings reports or product launches, especially over shorter timeframes where its predictive power is stronger. Similarly, financial analysts could integrate sentiment data with traditional market indicators to develop a more holistic view of stock performance. For AI practitioners, these results suggest that sentiment analysis could be integrated with other models that analyze factors sentiment analysis cannot, further enhancing the overall effectiveness of using AI within stock predictions.

Based on this study, while sentiment analysis can be valuable in certain instances across different fields, it alone is not sufficient in predicting stock movements accurately. A more effective approach would likely involve combining sentiment analysis with more complex models that are trained on a broader set of economic data and take into account a wider range of influencing factors. By integrating sentiment analysis with these advanced algorithms and a more diverse set of news sources, a more accurate model could be developed, potentially improving the predictive power of such methods in the future.

## References

- 1 J. Shen and M. O. Shafiq, *Journal of Big Data*, 2020, **7**, year.
- 2 A. Mittal and A. Goel, *Stock prediction using Twitter sentiment analysis*, Stanford university, cs229 technical report, 2012.
- 3 B. Fazlija and P. Harder, *Mathematics*, 2022, **10**, year.
- 4 N. S. A. C. Repository, *Sentiment analysis implementation for New York Times articles*, 2024, [https://github.com/theguyisgood1/NYT\\_sentiment\\_analysis.git](https://github.com/theguyisgood1/NYT_sentiment_analysis.git).
- 5 I. M. Fund, *World Economic Outlook, April 2024*, International Monetary Fund, Washington, DC, 2024.
- 6 *SPDR S&P 500 ETF sentiment and price data*, 2024, <https://drive.google.com/file/d/1BCUBVPwDMiHmSU17gxDRDirgVdPjX6Qt/view?usp=sharing>.
- 7 *Sentiment Table 2*, 2024, <https://drive.google.com/file/d/1NmK7mwE5ex2COa5g0C4WMexMgWhioikh/view?usp=sharing>.
- 8 *Sentiment Data Table 3*, 2024, <https://drive.google.com/file/d/10qbrI1jPEoGA49CoDCpvNZTy6iVS1yhV/view?usp=sharing>.
- 9 *Sentiment Data Table 4*, 2024, [https://drive.google.com/file/d/1LF3Wpx-4KnoYECSS5do0s5MPv\\_Azplmu/view?usp=sharing](https://drive.google.com/file/d/1LF3Wpx-4KnoYECSS5do0s5MPv_Azplmu/view?usp=sharing).
- 10 *Sentiment Data Table 5*, 2024, <https://drive.google.com/file/d/16VqKikYAkx-wCSwxxRQHb8bYIjUalwRh/view?usp=sharing>.
- 11 *Sentiment Data Table 6*, 2024, [https://drive.google.com/file/d/1kjmlpOIH-5FARLYEW87VbM4bH2\\_YICal/view?usp=sharing](https://drive.google.com/file/d/1kjmlpOIH-5FARLYEW87VbM4bH2_YICal/view?usp=sharing).
- 12 *Sentiment Data Table 7*, 2024, <https://drive.google.com/file/d/14kgGeURUSLBQ84zIxGq6tPd1Y8hYE0N9/view?usp=sharing>.
- 13 *Sentiment Data Table 8*, 2024, <https://drive.google.com/file/d/1W2QuGQ3nnSk1FPbhxO2KQTCUHFYIYQHCr/view?usp=sharing>.