

CDS-492: Capstone for Data Science  
Fall 2025  
Presentation

Eric Wu

November 24, 2025

## **Topic:**

How powerful can a man's words be?

During his second term, the 47th president of the United States, Donald Trump, has inflicted chaos and pandemonium on the markets via his enlightened posts on social media. This unease has caused retail and institutional investors to second-guess their risk tolerance, long-term holdings, and confidence in the continued financial growth of the United States.

Do Donald Trump's colorful social media posts actually effect market performance in a measurable and meaningful manner the short and long-term?

## **Significance:**

The stock market affects all of us indirectly and directly for those with exposure in said markets. Some of us depend on continued growth in the stock market for our savings or retirement, obtaining stable employment from publicly traded companies who raise capital from healthy stock markets, or rely on the growing stock market as an economic signal that the economy is thriving, bolstering consumer confidence, which in turn boosts consumer spending.

## Prior research:

- **Market Sentiment:** Sentiment betas are key for understanding stock market movement [3] [8].
- **Data Sources:** Social media, online news, and forums are valuable sources for sentiment data [6] [9] [4] [5].
- **Techniques:**
  - Dictionary/Lexicon-based models (e.g., VADER) are performant on news and social media posts [6] [4].
  - Topic modeling and joint sentiment-topic models are popular predictors of stock price movement [7].

## **Deliverables:**

- Capstone project paper documenting my methodology, research, and findings
- Science paper documenting my findings
- Slide presentation for presenting my research and findings

## Method of inquiry:

I will be using the NADAQ and S&P500 indices' historic quotes from Yahoo Finance to approximate the overall effect of Donald Trump's social media posts on the broader market for ease of data compilation and a dataset of all of DJT's posts on Truth Social since his second term.



## Method of data collection:

The first two datasets contain stock movement data for the following index funds, the SP500, Standards and Poors 500, and IXIC, the NASDAQ 500. Yahoo fiance partners with trading exchanges and providers of financial market data to collect and provide historical stock quote data [2] [1].

### Yahoo Finance Historical Quotes for the NASDAQ and S&P500:

<i>Column</i>	<i>Comment</i>
Date	Date
Open	Opening price of the equity
High	Maximum price of the equity for the given day
Low	Minimum price of the equity for the given day
Close	The closing price of the equity

## Method of data collection (cont.):

The third data set used for this project is a web scrape of all of Donald J. Trump's posts on his social media platform, Truth social. This data includes his direct posts, re-posts, and his interactions with various truth social users. This dataset contains the following features:

<i>Column</i>	<i>Comment</i>
id	The unique identifier for the post
created at	Timestamp when the post was made
content	The text content of the post
url	Direct link to the post on Truth Social
media	An array of image and video URLs if the post contains media
replies count	Number of replies to Trump post
reblogs count	Number of re-posts, or re-truths, to Trump post
favourites count	Number of favorites to Trump post



## Theory:

The original hypothesis was that the sentiment of Donald Trump's posts would effect the stock market. I used index funds to abstract the stock market, in this case the S&P500 and NASDAQ Composite. My null hypothesis for this initial theory was:

$$H_0 : \mu_{positive} = \mu_{neutral} = \mu_{negative} \quad (1)$$

My second hypothesis was that the topic of Donald Trump's posts would effect the stock market. I used index funds to abstract the stock market, in this case, I only used the S&P500 index, because the performance of both indexes was similar over the timeframe of Trump's second presidency.

$$H_0 : \mu_{topic_i} = \mu_{topic_{i+1}} = \mu_{topic_n} \quad (2)$$

## Capstone:

The project integrates what I've learned in:

- CDS-230 Modeling and Simulation I
- CDS-301 Scientific Information and Data Visualization
- CDS-321 Natural Language Processing
- CDS-402 Machine Learning

With the intention of discovering the power an individual's words can have.

## Results: Sentiment Analysis Findings

**Objective:** Determine if the sentiment of DJT's posts affects stock market volatility (S&P500 and NASDAQ).

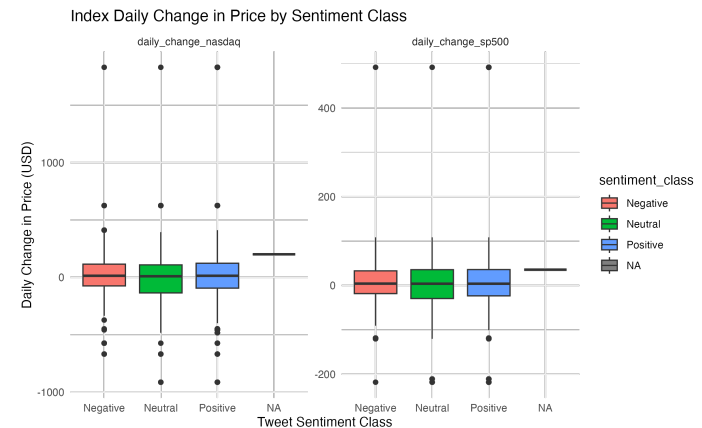
**Null Hypothesis ( $H_0$ ):** Mean volatility is equal across sentiment classes.

$$H_0 : \mu_{positive} = \mu_{neutral} = \mu_{negative}$$

### Findings:

- S&P500 p-value: 0.5216
- NASDAQ p-value: 0.4274

**Conclusion:** We fail to reject the null hypothesis. No statistically significant difference was found across sentiment classes.



## Results: Topic Modeling Findings

**Objective:** Determine if the topic of DJT's posts affects stock market volatility (S&P500).

**Null Hypothesis ( $H_0$ ):** Mean volatility is equal across topic classes.

$$H_0 : \mu_{topic_i} = \mu_{topic_{i+1}} = \mu_{topic_n}$$

### Single Anova Findings:

- S&P500 p-value: 0.000272

**Conclusion:** We find evidence to reject the null hypothesis. A statistically significant difference was found between topics.

	diff	lwr	upr	pval
3-2	28.20	0.82	55.59	0.04
7-3	-33.48	-58.98	-7.98	0.00
8-3	-25.02	-49.28	-0.75	0.04
9-3	-33.83	-62.33	-5.33	0.01

### Post-Hoc Tukey LDA

<sup>a</sup>

<sup>a</sup>The topics are described as:  
 Topic 2: Negative posts;  
 Topic 3: Trade War; Topic 7: Positive posts;  
 Topic 8: War; Topic 9: Border control/immigration

# Results: seededLDA Topic Modeling Findings

**Objective:** Determine if the topic of DJT's posts affects stock market volatility (S&P500).

**Null Hypothesis ( $H_0$ ):** Mean volatility is equal across topic classes.

$$H_0 : \mu_{topic_i} = \mu_{topic_{i+1}} = \mu_{topic_n}$$

## Single Anova Findings:

- S&P500 p-value: 0.00528

**Conclusion:** We find evidence to reject the null hypothesis. A statistically significant difference was found between topics.

	diff	lwr	upr	pval
TRADE-other1	24.99	4.34	45.64	0.01
TRADE-other2	23.11	1.22	45.01	0.03
TRADE-other3	21.42	0.09	42.76	0.05
WAR-TRADE	-25.48	-46.77	-4.18	0.01

## Post-Hoc Tukey seededLDA <sup>a</sup>

<sup>a</sup>The topics are described as:  
 Trade: Posts about tariffs and trade wars; Other1: Negative Posts; Other2: Positive Nationalism; Other3: Digs at Democrats; War: Posts about actual war

## **Conclusions:**

The tests I conducted on the use of sentiment to predict market movement were unexpectedly unfruitful. There were previous studies that found correlations between sentiment categories extracted from forms of textual media, like articles and social media posts about the markets, however, due to my singular focus on an individual's posts on an arguably smaller and potentially more niche social media platform and due to the reliance on pretrained and precomputed lexicon based sentiment analyzers, I was unable to reproduce the previous studies' results.

This inability to reproduce the previous studies's results could also be down to a small dataset of about 3000 useable texts, of which were around 200 words or less. Another potential error may be due to the VADER lexicon having specifically been trained on written text compared to the tendency of social media users to type out what is more similar to what they may say in conversation.

## **Conclusions (cont.)**

The lack of differences in the means of the daily volatility of both stock indexes, given the sentiment class, could be due to the VADER lexicon not performing well on my textual dataset, but it could also be due to the markets being more interested in the substance of specific DJT social media posts rather than the sentiment.

These results from the topic models on the influence of the topic of DJT's social media posts, the substantive content of the posts, on the market were found to be measurable. Unfortunately I was unable to chain together the influence of the sentiment and the topic. If there were to be more research on the subject of this research, I would recommend compiling a custom lexicon based sentiment analyzer based on DJT and other politicians or influential public figure's social media posts and their stated sentiments. I would also recommend exploring other topic models like top2vec or text embedding extraction.

## Source Materials:

- <https://www.marketwatch.com/investing/index/spx/download-data?startDate=1/1/2025&endDate=09/03/2025>
- [https://www.marketwatch.com/investing/index/comp/download-data?mod=mw\\_quote\\_tab](https://www.marketwatch.com/investing/index/comp/download-data?mod=mw_quote_tab)
- <https://github.com/stiles/trump-truth-social-archive/tree/main?tab=readme-ov-file>



## References

- [1] Historical stock quote for GSPC, 2025.
- [2] Historical stock quote for IXIC, 2025.
- [3] M. Baker and J. Wurgler. Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2):129–152, Spring 2007.
- [4] N. Das, B. Sadhukhan, T. Chatterjee, and S. Chakrabarti. Effect of public sentiment on stock market movement prediction during the COVID-19 outbreak. *Social Network Analysis and Mining*, 12(1):92, 2022.
- [5] C. Hutto and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):216–225, May 2014.
- [6] D. L. John, B. Stantic, T. K. Tran, U. Tukayev, T.-P. Hong, B. Trawiński, E. Szczerbicki, and N. T. Nguyen. Machine learning or lexicon based sentiment analysis techniques on social media posts. In *Intelligent Information and Database Systems*, volume 13758 of *Lecture Notes in Computer Science*, pages 3–12. Springer, Switzerland, 2022.
- [7] T. H. Nguyen, K. Shirai, and J. Velcin. Sentiment analysis on social media for stock movement prediction. *Expert systems with applications*, 42(24):9603–9611, 2015.
- [8] M. Nofer. The value of social media for predicting stock returns : Preconditions, instruments and performance analysis, 2015.
- [9] D. Shah, H. Isah, and F. Zulkernine. Predicting the effects of news sentiments on the stock market, 2018.