

Group: **Terence Johnson Fan Club**

DS 3001

Terence Johnson

31 March 2025

What is an observation in your study?

- An observation in our study will be each stock's daily start price, close price, high price for the day, low price for the day, volume of trades, the stock ticker symbol, and date.

Are you doing supervised or unsupervised learning? Classification or regression?

- We are doing supervised learning using a neural network in python to train our model. We are performing regression on the dataset to predict how key political events during Donald Trump's presidency impact the stock market.

What models or algorithms do you plan to use in your analysis? How?

- Initially, we will use linear regression to establish baseline relationships between political events and financial changes. We may also use LASSO regression to handle potential multicollinearity and reduce overfitting given the number of event dummies. If the relationships appear nonlinear or interactions are important, we may expand to decision trees or random forests. Time series models like ARIMA may be considered for trend analysis and stationarity checks.

How will you know if your approach "works"? What does success mean?

- Success means identifying statistically significant and interpretable relationships between specific political events and financial market reactions. We'll assess model performance using metrics like R^2 , RMSE, or mean absolute error for regression tasks. If using classification, we'll evaluate accuracy, F1-score, and confusion matrices. Additionally, clear visual patterns in time series plots before and after events would reinforce our findings.

What are weaknesses that you anticipate being an issue? How will you deal with them if they come up? If your approach fails, what might you learn from this unfortunate outcome?

- One major challenge is identifying and accurately coding political events, which introduces subjectivity. We'll mitigate this by sourcing events from credible, consistent outlets and documenting our criteria. Another issue is timing lag as markets may react with delay or anticipation. We will explore lagged variables and event windows (e.g., ± 3 days). If we find no strong relationships, we may conclude that market reactions are more nuanced or driven by broader economic factors—which is still a valuable insight.

Feature Engineering: How will you prepare the data specifically for your analysis? For example, are there many variables that should be one-hot encoded? Do you have many correlated numeric variables, for which PCA might be a useful tool?

- Within our data we plan to have a number of one-hot encoded variables about President Trump's activity. Dummy variables could include press conferences, executive orders, tweets about tariffs, major news articles being published regarding the administration's dealings with any of our sectors, etc. The rest of the data would consist of the columns containing the stock ticker, the daily adjusted closing stock price, and which sector the stock belongs to. The sector column will also be hot encoding in order to perform PCA to uncover trends within each sector.

Results: How will you communicate or present your results?

- We think displaying the results through neural networks is the most efficient way to show the effect the different types of political events are having on stock fluctuations. We hope this modeling will show a clear positive relationship between the two variables.

We will either use a measure of political activity (executive order specifically) or multiple biases.