Analyzing Recent Donald Trump Articles in North America

Ellis Guo , Team member 1 , Team member 2

[1] McGill University

ellis.guo@mail.mcgill.ca,

Introduction

When American businessman and reality-TV star Donald Trump launched his political career by announcing his candidacy for President of the United States in 2015, he quickly became one of the most polarizing and widely covered figures in modern media (Duignan 2024). Serving as the 45th President from 2017 to 2021, Trump has maintained significant public and media attention since leaving office (Duignan 2024). As he campaigns for the 2024 presidential election, Trump's influence continues to dominate headlines across North America (Duignan 2024).

These recent developments have resulted in substantial media coverage. We were hired by a media company to analyze how North American media portrays Donald Trump and the topics that dominate coverage of him. **Our report examines recent, English-language news articles from a range of North American news outlets, focusing on what aspects of Trump's life and actions are being covered and whether this coverage leans positive, negative, or neutral**. Relevant stakeholders include our client, as well as potentially Trump's political team or public figures interested in understanding media trends related to political figures. For example, stakeholders may seek insight into how specific events, such as campaign rallies or legal challenges, shape public portrayal.

With these stakeholders in mind, we define the scope of the analysis as follows:

"Donald Trump coverage" refers to any articles that mention his name in the title or description. This includes coverage of his political campaigns, personal life, legal issues, or references to related figures (e.g., family members, political allies). Since public perception of political figures is influenced by the actions of their associates and critics, this broader scope captures the context surrounding Trump's media portrayal.

"Recent" defines a timeframe of approximately two months during October 2024 and November 2024, providing a longitudinal view of Trump-related media coverage leading up to and after the election.

"North American news outlets" are defined as official news sources based in the United States or Canada with an online presence. The analysis is limited to news media, excluding social platforms such as X (formerly Twitter), Facebook, or Instagram.

After analyzing some 507 articles following the specifications above, we present a set of key findings.

1. Recent media coverage of Donald Trump tends to be predominantly neutral or negative in tone.

2. The most frequently addressed categories in the articles are related to election details and Trump's cultural impact.

3. Economic discussions during the analyzed period were largely centered on topics such as tariffs and cryptocurrency.

4. Coverage of Trump's criminal case has been notably negative and has garnered substantial attention compared to other facets of his career.

Data

Our dataset records key data on 507 articles. The dataset includes the following columns defined as followed:

- *Title*: The title of the news article.

- *Description*: The subtitle or, if unavailable, the first sentence of the article's body.

- *Source*: The name of the news provider that published the article.

- *Category*: The inferred category of the article's content (e.g., "election"), based on human annotation.

- *P/N/Ne*: The inferred sentiment of the article, classified as Positive, Negative, or Neutral, based on human annotation.

The data was collected using the NewsAPI, which imposed several constraints that influenced the collection design:

1. **The NewsAPI free plan restricts the lookback period to 30 days.** To overcome this limitation, data was collected in multiple rounds, each covering a 30-day window. This restriction led to a focus on "recent" articles about Donald Trump, specifically two weeks before and after the American election, which was the central event of the analysis.

2. **The API limits the number of results per request to 100 articles.** To collect enough data, 100 articles were collected from five time periods of five to six days from October 16th,2024, to November 14th, 2024.

3. **The API does not inherently insure a diverse selection of sources.** Initially, the dataset had a significant over representation of three sources out of ten individual sources. To overcome this, we identified 54 available North American sources from NewsAPI and individually selected articles from each source. The final dataset had 41 sources that yielded relevant results, ensuring a broader and more representative selection of sources in the final dataset.

To ensure fairness and avoid overrepresentation,the maximum number of articles per source was capped at 20. Duplicate articles were removed, reducing the dataset to 503 articles.

During manual annotation, 36 articles were identified and excluded because they either did not mention Donald Trump or were not focused on him, despite appearing in the title or description. To ensure full coverage, an additional 40 articles from the period November 15th, 2024, to November 17th, 2024, were collected, as no articles had been retrieved during this time in the original rounds. These articles were carefully checked to confirm their relevance to the subject of Donald Trump.

The final dataset consists of 507 articles, categorized into the following groups by their source:

- *Left-Leaning:* 103 articles
- *Right-Leaning:* 105 articles
- *Centrist or Neutral:* 151 articles
- *Entertainment/Technology/Other:* 67 articles

Methods

We explain methodologies involved in the following: development of the typology, tf-idf analysis, further analysis.

Typology: We annotated two inferred features: *category* and *P/N/Ne* (see "data"). Starting with open coding of 200 articles, we refined the typology through two rounds of coding. The first round identified 12 categories, which were then merged into a final set of 6 categories after further discussion. This process eliminated the need for an "other" category, as all articles fit comfortably within the defined categories.

Tf-idf: We created an algorithm to calculate word frequencies in each of the six categories. To focus on meaningful terms, we removed stop words like "the" and "and," as well as "Donald" and "Trump," since these words appeared too often to provide useful insights. We kept the top 10 words for each category, including any words with the same frequency as the tenth word, ensuring that the selection was concise and relevant. Using these frequencies, we calculated tf-idf scores with the formula from class. We visualized the results using bar plots created with the matplotlib library and pie charts in Excel, making the key terms in each category easier to understand.

Further Analysis: After the tf-idf analysis, we analyzed sentiment by grouping articles into positive, negative, and neutral categories for each typology. Articles were classified as follows:

- *Positive:* Articles that portrayed Trump in a favorable light or expressed support.
- *Negative:* Articles that criticized Trump or created a negative impression through unfavorable commentary.
- *Neutral:* Articles that presented factual, unbiased information without opinion.

We calculated the proportions of positive, negative, and neutral articles in each category and visualized these results using pie and bar charts generated with matplotlib and Excel respectively. This approach made it easier to see sentiment trends. We also examined the distribution of articles by news provider.

Results

From analyzing the various articles about Trump, we created six different categories that perfectly sum up the content that the articles fall in to.

1. **Election:** This category contains articles that speak of the Trump vs. Harris election. This category includes anything relating to campaigning/debates/interviews/statements for presidential candidates, articles that mention an endorsement, or articles that mention the result of the election.

   - Examples:
   – *"Donald Trump elected U.S. president in historic comeback"*: This one is relating to the results of the election
   – *"Microphone malfunction: Trump waits on stage for 18 minutes during 'technical difficulties'"*: This is due to it relating to Trump campaigning

2. **Polling:** This category refers to any articles that are related to the forecasting of the electoral college. This includes anything relating to polling and percentages for either candidate, swing states, voter demographics, or articles that are focusing on a state and the impact it has on the election.

   - Examples:
   – *"Trump will win Iowa, CBS News projects"*: Talks about polling of a state
   – *"Some voters aren't crazy about Trump. But they're supporting him anyway"*: This one is because it's referring to voters

3. **Legal Issues:** This category refers to any articles that mention any legal and security issues. This includes anything relating to cases filed/being filed against Trump, judges or prosecutors' opinions on Trump, violence, or mentions of January 6th and the insurrection.

   - Example:
   – *"Trump has won the election. What happens to his legal cases?"*: Mentions Trump's legal cases

4. **Policies:** This category refers to articles that mention political and policy impacts. Essentially, articles that mention policies that either candidate is planning to invoke, or business protection policies.

- Examples:
  - *"Donald Trump Says States Will Undo Abortion Laws That Are Too Tough"*: Talks about the policies that will be invoked if he won

5. **Economy:** This category contains articles that mention financial instruments. So, anything mentioning the economy, market, or cryptocurrency.

  - Examples:
    - *"What does Trump's victory mean for Canada? All roads lead to a weaker loonie"*: Mentions the economy

6. **Cultural Impact:** This category refers to articles that mention the social or cultural impact of the election. This includes celebrity/social medias reactions to the election or Trump, general reactions on the campaigns/candidates/results, and the impacts on peoples/places

  - Examples:
    - *"Brian Cox Slams Donald Trump: 'Monster'"*: An opinion on Trump
    - *"President-elect Donald Trump has vowed to carry out what he has called the largest deportation effort in American history. Here's how those on both sides of the Southern border are reacting to his election and claims."*: This article mentions how people are reacting to the election and to Trump
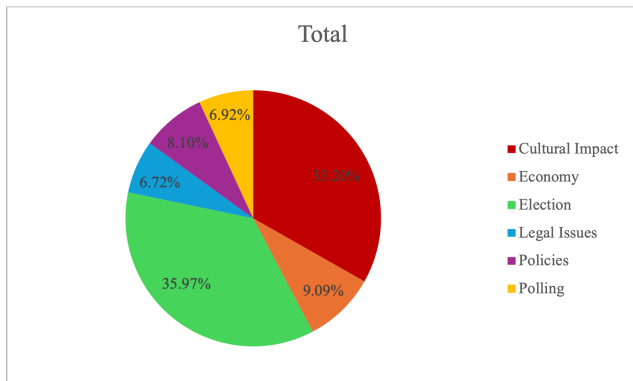

Figure 1: Breakdown of Percentage of Articles per Category.

The second feature we used as a form of analyzing, is a ranking of whether the articles are "negative", "neutral", or "positive" per category. Most of the categories are pretty evenly spread between negative and neutral, with positive being the minority in all of the categories. The outlier is the category "Legal Issues" and to a lesser extent "Election", where there are more neutral articles than negative.

Figure 2: Pie Charts of P/N/Ne (Sentiment) of Article per Category.

Additionally, we found out the top 10 tf-idf words for each category. When you analyse the graphs, you can see that the category "Election" has the highest scores, which indicates it has the most distinctive vocabulary. In the categories "Polling" and "Legal Issues" there are clusters of the same score, which indicates that these terms appeared in a similar frequency to each other. Most of the categories are uniform, except for the category "Election", which is skewed.
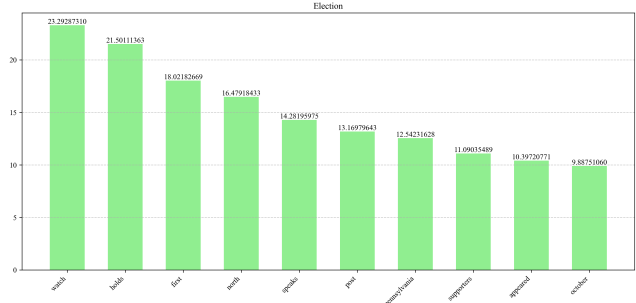

Figure 3: Bar Plot Visualizing Tf-Idf Frequencies of Election.
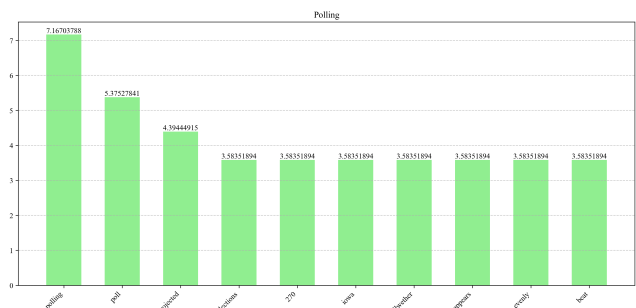

Figure 4: Bar Plot Visualizing Tf-Idf Frequencies of Polling.
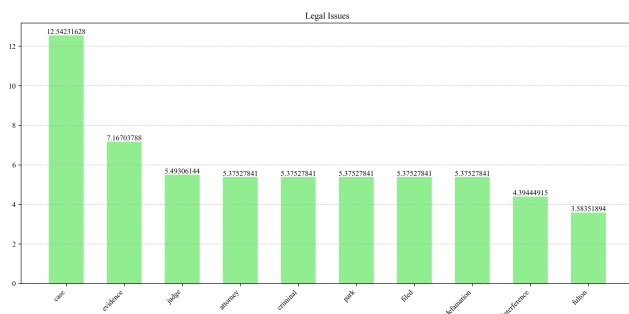
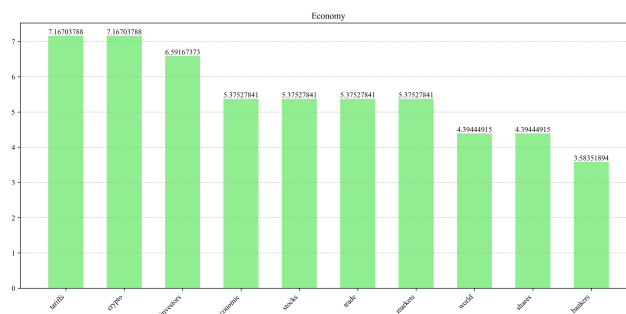Figure 5: Bar Plot Visualizing Tf-Idf Frequencies of Legal Issues.

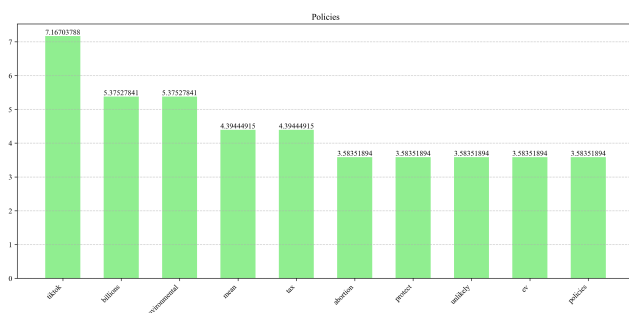Figure 7: Bar Plot Visualizing Tf-Idf Frequencies of Economy.

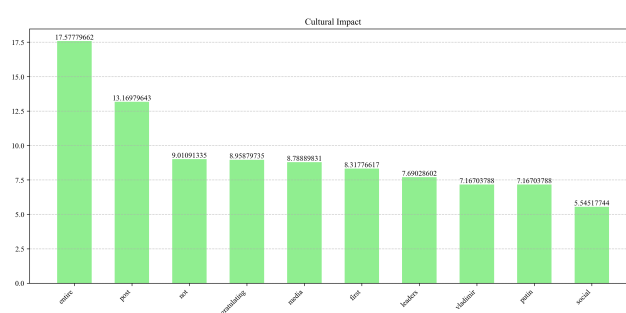Figure 6: Bar Plot Visualizing Tf-Idf Frequencies of Policies.

Figure 8: Bar Plot Visualizing Tf-Idf Frequencies of Cultural Impact.

Discussion

Of the six categories, the vast majority of articles belong to either the "Election" or "Cultural Impact" category, with both together accounting for roughly 70% of all articles, while the other 4 categories together accounted for the remaining 30% of articles. This makes sense, as with the articles being collected during the time-frame of October and November 2024, they are perfectly aligned

with the timing of the US presidential elections featuring Donald Trump and Kamala Harris. It is by far the biggest and most important event happening in regards to Donald Trump as the outcome of the election has significant impacts everywhere throughout North America and even the rest of the World. In fact, all other categories are also in some way related to the election; "Polling", "Policies", "Economy" and "Cultural Impact" all cover the consequences of Trump either winning or losing.

The "Election" category's top 3 scoring tf-idf words are "watch", "holds" and "first", reinforcing the importance of the election and the need for audiences to closely monitor every step of its process and outcome. It can be noted that, alongside "Legal Issues", "Election" has a predominantly neutral coverage (over 50), followed by a large portion of articles having negative coverage, and finally a few articles with positive coverage. This can be seen as news outlets wanting to give objective information regarding the election regardless of their personal biases, favoring factual coverage of events without incorporating too many personal opinions.

The same is the case with the "Legal Issues" category, having the largest (60+%) amount of neutral coverage, with the high scoring tf-idf words being "case", "evidence", "judge", "attorney", "defamation" and "interference" all reflecting the multitude of legal cases facing Trump. Again, the favoring of neutral coverage can be observed for this category; the topic is very serious, especially for a presidential candidate, and the need for factual information is crucial. The words "fulton" and "criminal" further reinforce the seriousness of the situation, as a mug shot of Trump was taken after he was arrested and brought to Fulton County Jail.

As for the other four categories, for "Polling", the 3 highest scoring tf-idf words are "polling", "poll", and "projected", alongside other words like "iowa", "evenly" and "270". This reflects the predictions media outlets were making on the elections, with Iowa being a surprise swing state and polls for both Trump and Harris being very close. The number 270 is here as it is the exact minimum number of electoral votes required to win the elections. "Polling" interestingly has the largest proportion of positive coverage (22%) compared to all other categories (7-15%) while also being the category furthest from Trump himself, relating more to predicting the outcome of the election rather than coverage of his person. It is also the only category that, once the election ends, will have no further major impact other than confirming or rejecting the forecasted results.

In regard to "Economy", words such as "crypto", "investors", "stocks", "trade", "markets" and "shares" appear

often when consulting the tf-idf scores, with "tariffs" and "world" also scoring highly, implying an emphasis on the economic impact the election will have on the stock market alongside effects on international trade.

Then, observing the "Policies" category, the tf-idf scores have the words "tiktok", "environmental", "abortion", "tax", and "protect". This category of articles suggests an interest in all fields that Trump's policies would cover, from his stance on social media to the environment to abortion and tax laws, all the way to the military.

Finally, the category "Cultural Impact" is harder to analyze. The high-scoring tf-idf words "entire", "post", "not", "first", "social", aren't clear in their meaning and further analysis would be needed to come to a conclusion. On the other hand, "congratulating", "media", "leaders", "vladimir", "putin" seem to suggest an interest in the responses of various world leaders (notably Vladimir Putin) and media outlets to the election, with congratulations being focused on. It is interesting to note, however, that despite that being the case, this category has the largest proportion of negative coverage, with 44% of articles being as such.

Overall, 40% of articles have negative coverage, 49% of articles are neutral, and 11% are positively covered. Other than "Election" and "Legal Issues", which have majorly neutral (50+%) coverage, all other categories had roughly as much negative coverage as neutral coverage (within 5% of each other) and hovered around the 40-45% range. Positive coverage on the other hand was minimal and varied from 7-22%. Overall, this suggests that news coverage surrounding Trump during the election period is mostly neutral or negative, with very little positive articles about him. However, it is important to recall that in our study, we defined articles as Neutral if they were factual about their information about Trump, Negative if they were biased in a bad way against Trump, and Positive if they were biased in a good way against Trump. So, if an article were to cover an objectively bad or good subject surrounding Trump, for example, Wall Street has already priced in a Trump victory—just look at European stocks and the inflation market would be classified as a Neutral article despite the violent subject. Our study shows that for the period surrounding the elections, news coverage on Trump were mostly factual or biased negatively towards him.

It is important to note the time frame of the articles collected was during the elections, which almost definitely played a role in the coverage trends observed in this study and, thus the trends here should not be used to generalize a different time frame (i.e. post-election, pre- /post- taking office, etc.). All in all, the analysis reveals a notable tendency towards neutral coverage, particularly in the "Election" and "Legal Issues" categories, where over 50% of articles were classified as neutral. However, while neutral reporting on pivotal topics like the election and legal cases was prevalent, the overall sentiment in other categories leaned towards negativity. Future studies could expand on our findings by exploring how these trends evolve over time or differ across various news outlets and platforms throughout different countries.

References

Duignan. 2024. Encyclopedia Britannica 2024. https: //www.britannica.com/biography/Donald-Trump.