

COMP 370 Final Project – Donald Trump’s Coverage in the North American Media

Written by

Duru Aran, Omar Assaadi, Leen Assaf

3480 Rue University, Montreal, QC H3A 2A7

duru.aran@mail.mcgill.ca, omar.assaadi@mail.mcgill.ca, leen.assaf@mail.mcgill.ca

1 INTRODUCTION

This report contains the analysis of over 500 North American news articles covering Donald Trump, collected in the weeks leading up to and following the recent US election, revealing many important insights into the media’s portrayal of Donald Trump. The analyzed data focused mainly on North American news coverage in English. The analysis and report’s primary focus is specifically on media sentiment and topic coverage. So, this report examines whether the media coverage was positive or negative and the topics covered the most.

The majority of articles analyzed conveyed an overall positive sentiment toward Trump, with less than a third categorized as having an overall negative sentiment. This indicates that, despite the polarized nature of political reporting, a significant portion of media coverage regarding Trump was in his favour.

When conducting the open coding and TF-IDF analysis, six key topics emerged as the main areas of discussion in the media. These topics are as follows: Consequences of Trump’s Re-Election, Election Analysis, Trump Campaign, Opinions on Trump and

his Supporters, Trump’s Policies, and Coverage Focusing on Kamala Harris.

The results showed that coverage was divided approximately evenly between the following four categories: Opinions on Trump and his supporters, Trump’s Campaign, Consequences of Trump’s Re-Election, and Election Analysis. The other two categories, namely Trump’s Policies and Coverage Focusing on Kamala Harris, only accounted for 5% of the dataset.

2 DATA

We started the data analysis project with a dataset containing 600 articles. The articles were parsed into a JSON file with the following columns, with the Initial Summary column left empty for the members to complete manually: Title, Description, Source, and Initial Summary.

This dataset was created using the free plan of the API newsapi.org. However, the free plan had some limitations that we must specify. The API has a maximum of 100 results per day. To get a sufficient number of articles, we collected data day by day, setting different timelines each time to avoid a large

number of duplicates. The free plan also returned articles that had only been published in the past 30 days, which could not be mitigated.

To get our dataset, we created a Python script that generated 59 news sources across North America (USA, Canada, Mexico) using the NewsAPI `‘/v2/sources’`. These sources were then used as input for another program in Python that used the NewsAPI `‘/v2/everything’` to retrieve the most recent 600 articles written in English from the selected North American sources.

To minimize bias, we limited 50 articles per source, which allowed us to maintain the diversity of sources in our dataset. Then, we performed basic data cleanup, which involved removing duplicates and missing values and ensuring the dataset was complete and ready for the annotation phase.

The group members conducted an open coding process on a subset of 200 articles to identify our topics. To ensure a fair representation from each source, these articles were randomly selected so that each news source would contribute the same percentage of articles as it does to the overall dataset. This was done to ensure the most accurate representation of the data during the open coding process. During annotation, any articles that were unanimously deemed irrelevant to Donald Trump by our group were removed. After the data cleanup was over, the resulting dataset contained 581 articles.

3 METHODS

During the open coding process, after sampling, each team member reviewed the articles, focusing mainly on the title and opening paragraphs, unless in times of uncertainty, and summarized them with a few words

that captured the essence of the article’s content. Then, we met to compare our open coding annotations, when we identified common themes and trends, which allowed us to cluster similar summaries into six distinct categories.

To better understand the key terms associated with each topic, a TF-IDF analysis was performed. Initially, we ran the analysis without removing stopwords, like “the”, “to”, or “in”, but we found that a majority of the top 10 words were generic stopwords. Then, we decided to re-run the analysis after removing stopwords to get a more meaningful representation. We extensively used Python, particularly the scikit-learn library’s “TfidfVectorizer” function, to conduct this TF-IDF analysis.

We tried several different tools for sentiment analysis but finally settled on using the VADER2 tool for the final analysis, as it provided a more nuanced understanding of sentiment compared to the tools we had tried, like TextBlob or BERT. VADER2 could distinguish between different sentiment levels, like “strongly positive” vs. “positive,” and allowed for a broader range of sentiment classifications. It also provided a more uniform distribution across the dataset, unlike the TextBlob tool, which tended to categorize a majority of the articles as “neutral”.

These methods provided a comprehensive approach to analyzing North American media coverage of Donald Trump. They combined identifying the topics with keyword and sentiment analysis to uncover key insights into the media’s portrayal of him during and after the elections.

4 RESULTS

Below are the final definitions of the six topics used during the annotation phase.

4.1 Topic Definitions

Consequences of Trump's Reelection – This category covers articles examining what happened during and after Trump's election victory. This category includes articles about his cabinet picks, media and public opinions on the consequences of his win, reactions and congratulations to Trump but also his response to re-election, and endorsements he received.

Opinions on Trump and his Supporters – This category focuses on exploring the different points of view regarding Trump and his supporters. Coverage includes articles about his family, former employees opposing Trump, conflicts between liberals and supporters, issues involving Chinese hackers, points of view from figures like Elon Musk, comparisons to historical figures like Hitler, opinions from Republican politicians, assassination attempts, Catholic viewpoints, and young people's views on Trump.

Trump Campaign – This category includes articles on various aspects of Donald Trump's campaign. Topics include articles about campaign-related violence, allegations of racism, Trump merchandise, campaign debts, campaign optimism, election corruption, and events at Trump rallies. Some articles might discuss financial challenges and accusations of misconduct during the campaign.

Trump's Policies – This category shows articles about Trump's policy positions. Articles topics include abortion rights, education plans, tax policies,

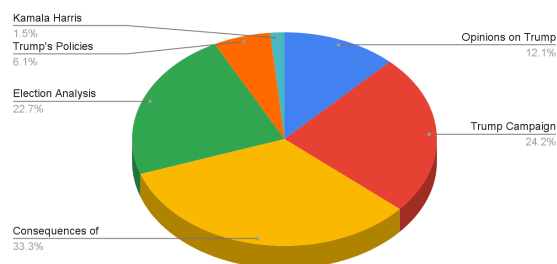
transgender policies, healthcare policies, and foreign policy. An important focus is put on the implications and public reactions to these policies.

Election Analysis – Articles in this category analyze the election process. Articles include discussions on campaign strategies, polling accuracy and analysis, and overall election coverage. These articles provide helpful information about how Trump achieved re-election.

Coverage Focusing on Kamala Harris – This category focuses on news about Kamala Harris. Articles include her distancing from Biden, reactions to election outcomes, support from the ultra-rich, her role in feminist empowerment, endorsements she receives, and events at her rallies.

4.2 Topic Distribution

Below are graphs that represent our findings, along with captions detailing what they entail.

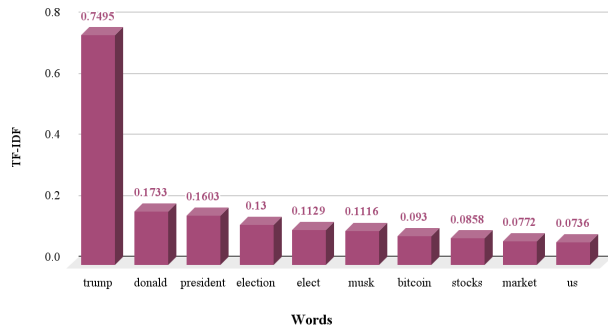


Graph 1 – Percentage Distribution of Categories

4.3 TF-IDF Results

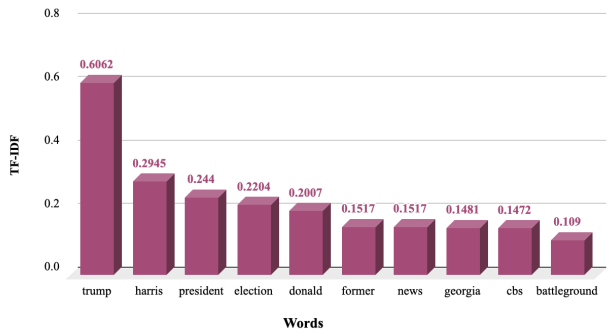
Below are the results of the TF-IDF analysis conducted on each category. The frequencies are given in percentages with respect to the overall frequency of only the Top 10 words.

Consequences of Trump's Re-Election



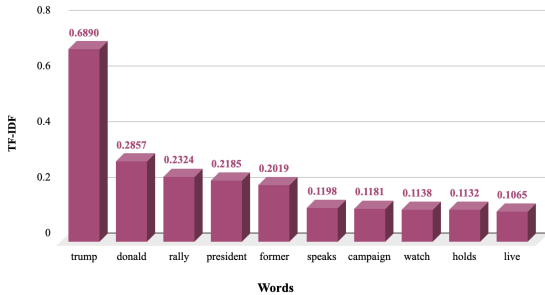
Graph 2 – TF-IDF Results of “Consequences of Trump’s Reelection”

Election Analysis



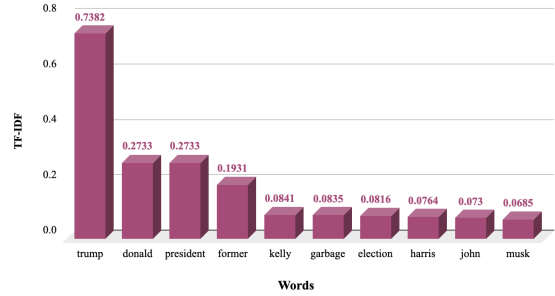
Graph 3 – TF-IDF Results of “Election Analysis”

Trump's Campaign



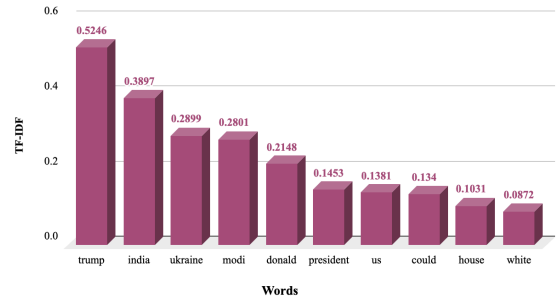
Graph 4 – TF-IDF Results of “Trump Campaign”

Opinions on Trump and his Supporters



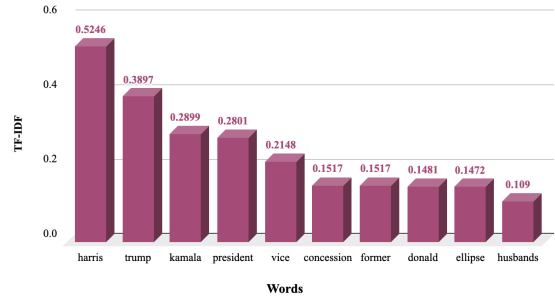
Graph 5 – TF-IDF Results of “Opinions on Trump and his Supporters”

Trump's Policies



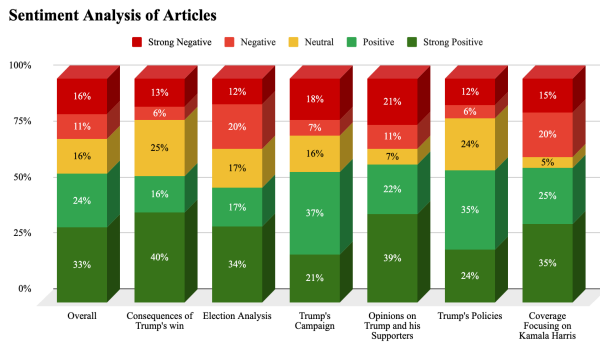
Graph 6 – TF-IDF Results of “Trump’s Policies”

Coverage Focusing on Kamala Harris



Graph 7 – TF-IDF Results of “Coverage Focusing on Kamala Harris”

4.4 Sentiment Analysis



Graph 8 – Sentiment Analysis Results, with respect to each topic.

5 DISCUSSION

From the results of sentiment analysis, it is evident that the majority of articles (over 50%) portray Trump positively, with a significant portion (27%) being classified as negative sentiment. This suggests a relatively polarized view of Donald Trump in North American media. The interpretation of results for each category's sentiment analysis is as follows.

5.1 Sentiment Analysis

Consequences of Trump's Re-Election – This category is the most positively skewed, with 56% of the articles in this category classified as “positive” (16%) or “strongly positive” (40%). It is also important to note that the percentage of “neutral” articles was higher than “negative” and “strongly” negative ones combined, implying that there were few articles that considered Trump's win to have negative consequences. This indicates that discussions on the aftermath of Trump's election, such as economic effects and cabinet picks, were largely optimistic.

Trump Campaign – This is another category that received highly positive sentiment, at 58% of the articles being perceived as “positive” (37%) and “strongly positive” (21%). However, there is also a notable portion of “strongly negative” sentiment at 18%, which could possibly be driven by controversies revolving around his campaign's strategy during the election period.

Election Analysis – This category had a relatively balanced sentiment distribution, apart from the mainly “strongly positive” skew (34%). This is to be expected, as this category analyzed election projections and results, that the winner of the election would receive a majority “strongly positive” sentiment about his position in the election. This also implies, however, that the media had “doubts” about his performance in the election, which could account for the 32% “negative” and “strongly negative” sentiments.

Opinions on Trump and Supporters – This category was one of the most polarized overall, with only 7% of articles being analyzed as “neutral”. This makes sense, considering that opinions are less likely to be “neutral” by definition. However, this makes the percentages of positive and negative sentiments much more meaningful. With 61% of the articles being categorized as “positive” or “strongly positive” compared to the 32% “negative” and “strongly negative”, it is apparent that the media tends to view Trump in a positive light.

Trump's Policies – This category saw a positively skewed, but still relatively neutral trend. The significant percentages of positive and neutral sentiment likely relates to media outlets reporting the policies in a factual and neutral manner, while also

talking about the support that surrounds his policies. This also implies that the public generally regarded his policies positively.

Coverage Focusing on Kamala Harris – As the category with the fewest articles, and while it showed a positive skew, a considerable percentage was still categorized as negative, which reflects the contentious political climate surrounding her role as Vice President. This category showed the lowest percentage of neutral articles, showing that the topic of Kamala Harris was controversial and polarizing around the elections. It is particularly interesting to also notice the similarities between the sentiment distributions of this category and Opinions on Trump. This could imply that while Trump won the election, overall media sentiment was relatively the same regarding Donald Trump and Kamala Harris as people.

5.2 TF-IDF Analysis

In addition to sentiment analysis, we used TF-IDF to identify the most relevant terms associated with each topic. Although words like “Trump”, “Donald”, “Harris”, “election”, and “president” were shared amongst all categories, it was interesting to look at the words that were unique to each category.

Consequences of Trump’s Reelection – This category had some interesting words in its Top 10 TF-IDF analysis, like “bitcoin”, “stocks”, and “market”, indicating that the main focus of the media when it came to how Trump’s reelection would affect the US was on the economy and the financial market. This is in line with Trump’s campaign that focused more on the economy and monetary policies.

Election Analysis – The words “battleground”, “georgia”, and “news” indicate that this category was mainly from election day and projections, as various news outlets were declaring and discussing election results. They were concerned with electoral analysis, particularly in key swing states, like Georgia.

Coverage Focusing on Kamala Harris – The Top 10 words in this category include “vice” and “husbands,” implying the media’s focus on Harris’s current power and personal life position. In contrast, the words “ellipse” and “concession” imply that most of the articles focusing on Kamala Harris but still including Donald Trump were about Harris’s loss of the election and her concession speech.

Opinions on Trump and Supporters – Including the word “garbage” in the Top 10 indicates a strongly negative sentiment towards Donald Trump and his supporters, which contrasts the results found in the sentiment analysis. However, the presence of words like “John”, “Musk”, and “Kelly” indicates that this category was mainly focused on specific people’s opinions on Donald Trump, which would likely indicate either strong support or strong opposition, leading to the small neutral percentage (7%).

Trump’s Policies – The fact that the only words that were not directly tied to President Trump were “Ukraine”, “India”, and “Modi”, indicates that the news mainly covered Trump’s foreign policy. The majority of positive sentiment for this topic also suggests that North American news outlets tended to support Trump’s foreign policy choices.

Trump Campaign – We didn’t find that any of the words in the Top 10 were particularly interesting, as they can all be tied to various different Trump Campaign strategies.

6 TEAM MEMBER CONTRIBUTIONS

Much of this report was written collectively, as a group. We coded the scripts, open coded, and annotated together, at the same time. Although everyone collaborated on the report as well, below are specific breakdowns of the parts each member worked on the most:

Duru Aran – Methods, Discussion, VADER2 Sentiment Analysis, Graphs.

Omar Assaadi – Introduction, Results, TextBlob Sentiment Analysis.

Leen Assaf – Data, TF-IDF Scoring Code, BERT Sentiment Analysis.

7 CITATIONS

Karabiber, F. TF-Idf - Term Frequency-Inverse Document Frequency. *Learn Data Science*, www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/.

Malde, R. 6 July 2020. A Short Introduction to Vader. *Towards Data Science*, towardsdatascience.com/an-short-introduction-to-vader-3f3860208d53.

pawangfg. 25 Sept. 2023. Sentiment Classification Using Bert. *GeeksforGeeks*, www.geeksforgeeks.org/sentiment-classification-using-bert/.

Watson, E. 13 May 2024. A Beginner's Guide to Performing Sentiment Analysis on Text with Python. *Medium*, medium.com/@eleanor.watson/a-beginners-guide-to-performing-sentiment-analysis-on-text-with-python-3ce80dcac22e.