An Analysis On The Political Landscape In Malaysia Based On Twitter

by Faiq Ali

Submission date: 21-Jun-2023 02:07PM (UTC+0800)
Submission ID: 2120162457
File name: WIX3002_Assignment_2__Group_2.pdf (1.18M)
Word count: 2771
Character count: 14982

An Analysis On The Political Landscape In Malaysia Based On Twitter

WIX3002 SOCIAL INFORMATICS

GROUP 2

ASSIGNMENT 2

Profesor Madya Dr. Kasturi Dewi A/p Varathan

21 JUNE 2023

Raja Zareef Firdaus Bin Raja Azman Nahar	17207394/2
Muhammad Waiz Wafiq bin Mohamad Jafri	17203410/2
Muhammad Amirul Hakiem bin Sabarudin	17202262/2
Syed Faiq Ali Faisal	S2013598
Ain Zaffry Bin Rizwal	17204733/2

Introduction

This analysis focuses on exploring the political landscape of Malaysia using Twitter data. By employing social network analysis, topic modeling, and sentiment analysis techniques, we aim to uncover the prevailing sentiments, popular topics, and influential individuals and communities within the Malaysian political discourse on Twitter. Through mapping user interactions, identifying key themes, and evaluating sentiment, we seek to gain valuable insights into public opinion and the dynamics of political discussions in the digital realm.

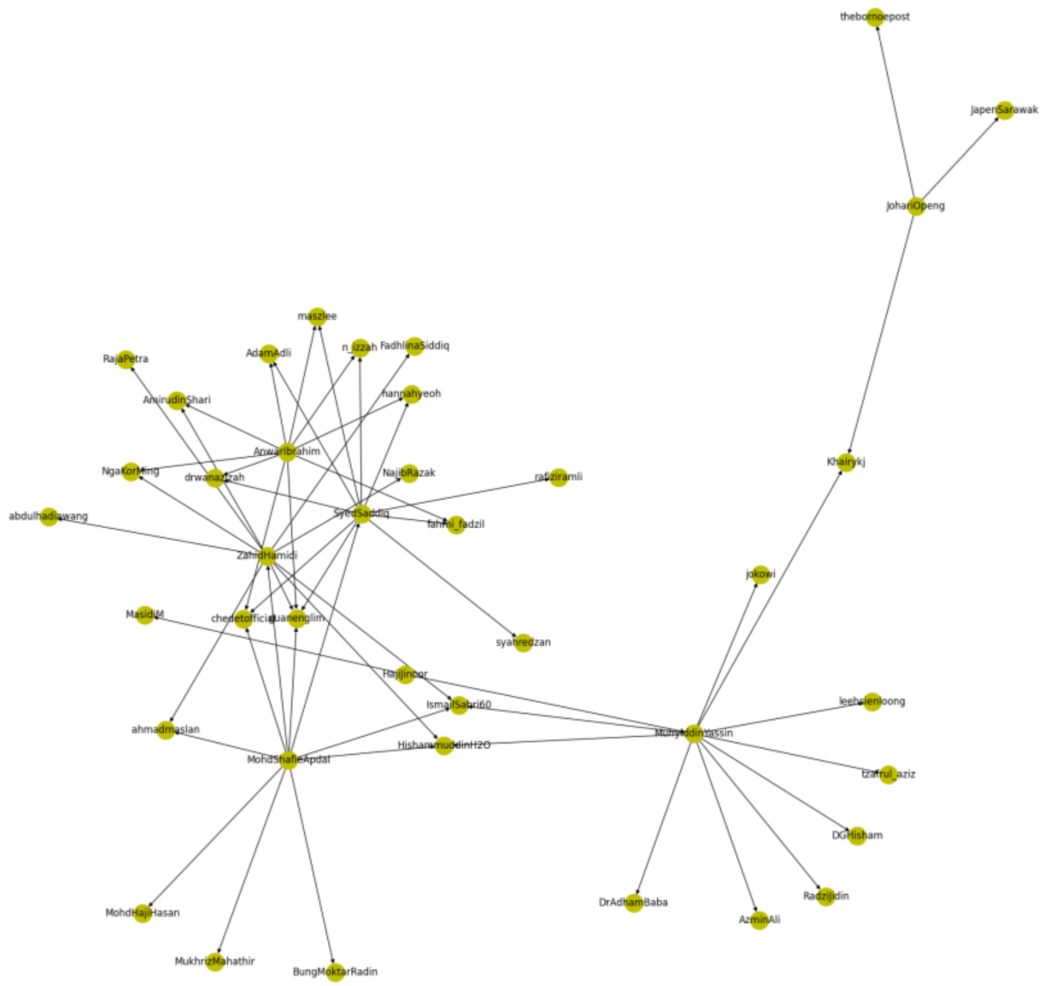
Social Network Analysis

Before constructing the network, we need data. We cannot include all political figures in Malaysia as it would involve hundreds or thousands of nodes & edges within the graph. In addition to that, not all politicians are active on social media. This is the limitation that we need to acknowledge within this study. We decided on studying the social dynamics of the leaders of coalitions that won seats in parliament, this includes the government & opposition, as they are:

Abbr	Leader	Leader's Twitter	Ideology	Position	Dewan Rakyat	Dewan Negara	State Assemblies	Vote share (2022)	Federal government
PH	Anwar Ibrahim	@anwaribrahim	Social democracy	Centre-left	82 / 222	14 / 70	162 / 607	37.46%	Government
PN	Muhyiddin Yassin	@MuhyiddinYassin	National conservatism	Centre-right to right-wing	74 / 222	15 / 70	150 / 607	30.35%	Opposition
BN	Ahmad Zahid Hami	@DrZahidHamidi	Conservatism	Right-wing	30 / 222	21 / 70	142 / 607	22.36%	Government
GPS	Abang Abdul Rahm	@AbgJohariOpeng	Sarawak nationalism	Centre-right	23 / 222	Jun-70	76 / 607	3.94%	Government
GRS	Hajiji Noor	@Hajijinoor	Sabah nationalism	Centre-right	6 / 222	Feb-70	42 / 607	2.98%	Government
Warisan	Shafie Apdal	@mohdshafieapdal	Sabah progressivism	Centre-left	3 / 222	0 / 70	14 / 607	1.82%	Government
KDM	Peter Anthony	N/A	Sabah regionalism	Centre	1 / 222	0 / 70	3 / 607	0.34%	Government
PBM	Larry Sng	N/A	Multiracialism	Centre-left	1 / 222	0 / 70	2 / 607	0.11%	Government
MUDA	Syed Saddiq	@SyedSaddiq	Populism	Centre-left	1 / 222	0 / 70	1 / 607	0.48%	Government

The aim of this social network analysis is to answer the following research questions:

Network Structure:
https://github.com/waizwafiq/wix3002_assignment



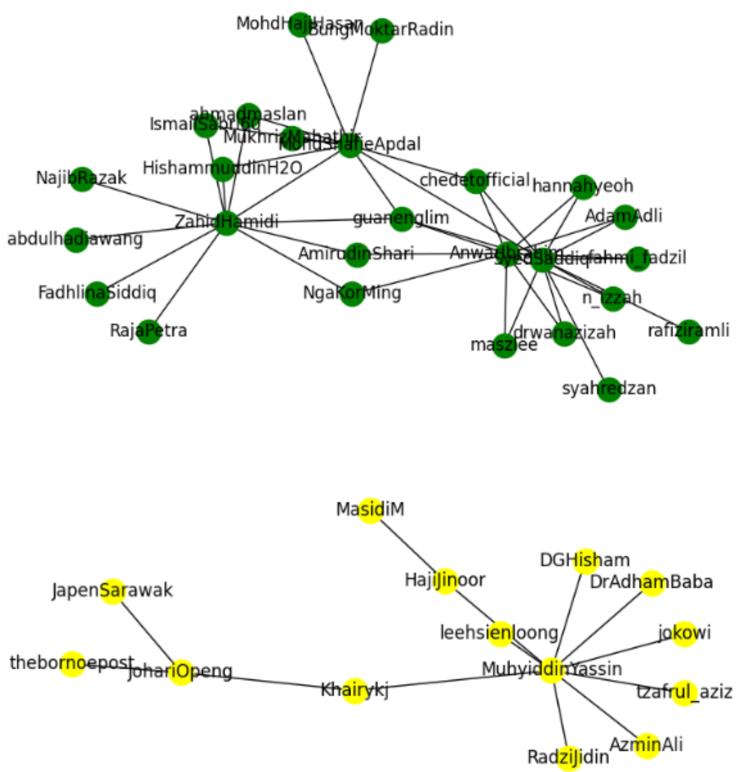
- What are the key network metrics, such as centrality measures (e.g., degree centrality, betweenness centrality) and network density?
  The highest degree centrality: 0.28205128205128205.

  Highest betweenness centrality is MuhyiddinYassin (0.0067476383265856945), followed by SyedSaddiq (0.005398110661268556) & ZahidHamidi (0.004048582995951417).

  Network density= 0.05128205128205128
  The graph has a density of 0.05 (5% of all possible edges present). This means that the graph is relatively sparse, with most nodes not connected to each other. This may be due to the relatively small number of samples taken as nodes for the graph.

Political Affiliations:

- Can we detect any patterns of clustering or homophily based on political ideologies or affiliations?
  Yes, by running the Girvan-Newman algorithm, where it uses edge betweenness centrality and iteratively removes edges with the highest betweenness, we are able to identify clusters within the network graph. Closer inspection on the nodes, we can see there are the “Government” (in green) & “Opposition” (in yellow) clusters, whereby the members of each cluster (the nodes) are homophilic in terms of political ideologies, especially apparent in the “government” cluster.



Influence and Information Flow:

- How does information flow within the network? Are there central individuals or groups that act as bridges or bottlenecks for information dissemination?
  In essence, closeness centrality indicates how easily information or influence can spread from a node to others in the network. Nodes with high closeness centrality are well-positioned to efficiently disseminate information or influence throughout the network. They facilitate communication or influence propagation.

These are some of the most central individuals based on the closeness centrality scores, in other words the best people to reach out to spread news faster:

1. guanenglim 0.10256410256410256
2. IsmailSabri60 0.08205128205128205
3. HishammuddinH2O 0.08205128205128205
4. chedetofficial 0.07692307692307693
5. Khairykj 0.057692307692307696
6. hannahyeoh 0.057692307692307696
7. n_izzah 0.057692307692307696
8. drwanazizah 0.057692307692307696

- Can we identify influential nodes based on their network centrality measures? How does their influence relate to their political positions or roles?
  In network analysis, centrality measures can help identify influential nodes within a network. In this case, we can examine the betweenness centrality and degree centrality of nodes to assess their influence in the political landscape.

  Betweenness centrality measures how often a node acts as a bridge or intermediary between other nodes in the network. Nodes with higher betweenness centrality have a greater potential to control the flow of information or resources. Therefore, they play a crucial role in the network's communication and information dissemination. In our network, Muhyiddin Yassin, Zahid Hamidi, and Syed Saddiq are identified as the most influential nodes based on their betweenness centrality scores of MuhyiddinYassin 0.0067476383265856945 ,0.004048582995951417, 0.005398110661268556 respectively. This indicates that these individuals have a significant impact on the flow of information within the network.

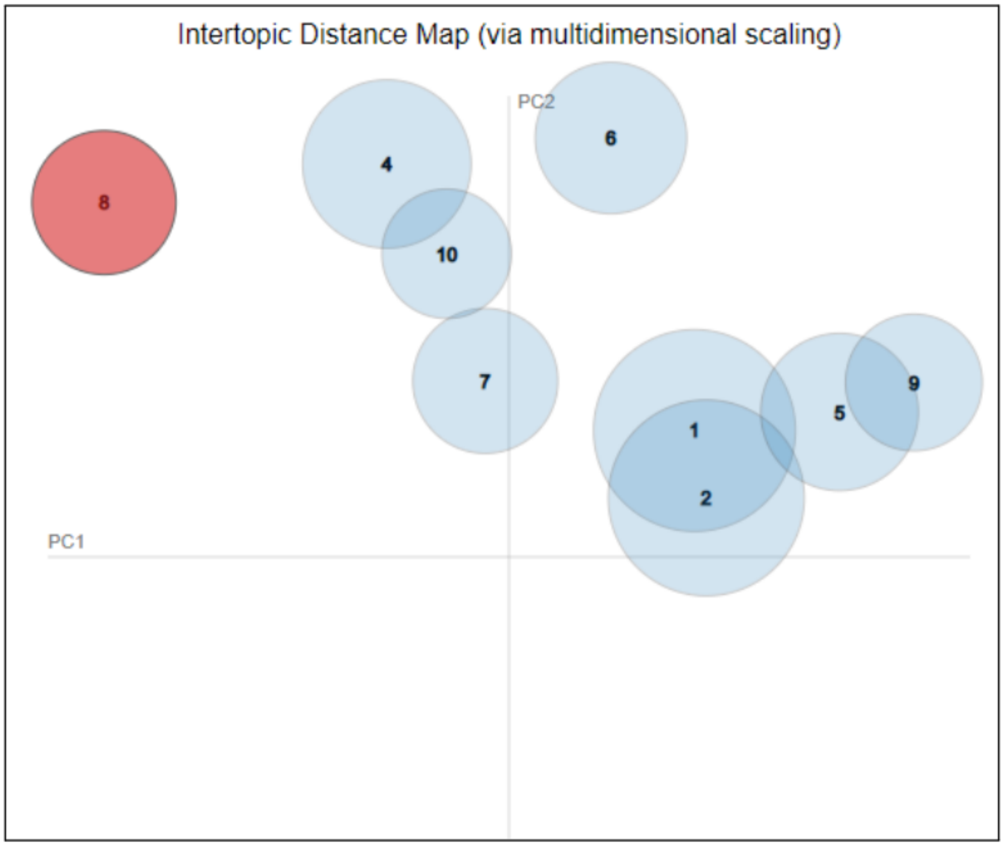
  Degree centrality, on the other hand, measures the number of connections a node has in the network. Nodes with high degree centrality have a large number of connections, which means they have a wide reach and can potentially influence a larger audience. In our network, the nodes mentioned earlier, Muhyiddin Yassin, Zahid Hamidi, and Syed Saddiq, have the same degree centrality score of 0.28205128205128205, suggesting that they are well-connected and have the ability to disseminate their ideas or messages to a substantial number of individuals.

Step 3: Topic Modeling

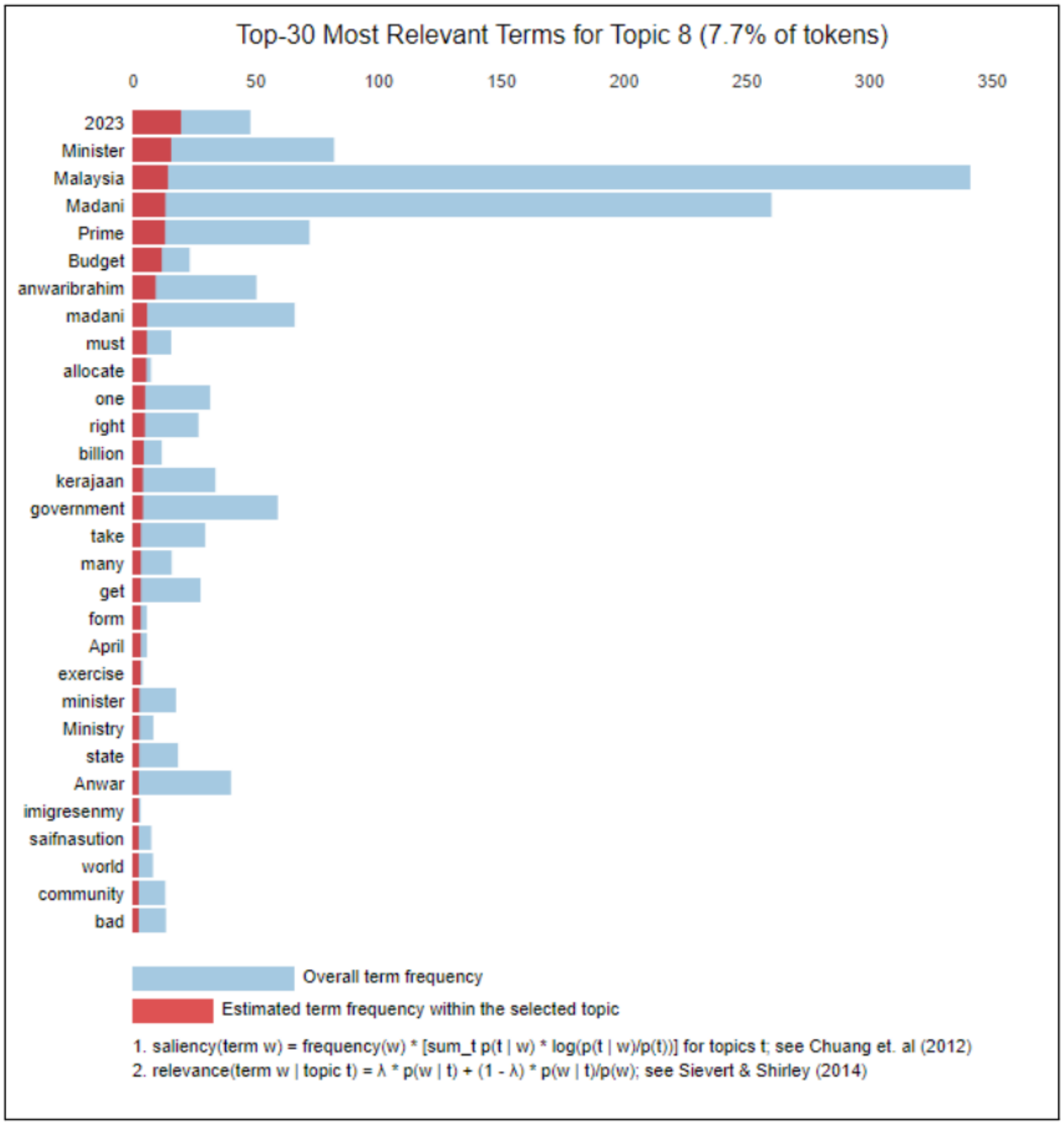
Our aim for this step is to pinpoint the trending topic concerning the current Malaysian government.

Query: KerajaanMadani

The justification for using this term stems from its historical context. This term, "KerajaanMadani," was introduced in December 2022 by Prime Minister Anwar Ibrahim. Since its inception, it has been widely adopted by Twitter users as a means to characterize the current Malaysian government in their political discussions on the platform. Consequently, it serves as an effective keyword to search for relevant discussions and trends related to the political scene of Malaysia on Twitter. By analysing the themes and topics related to this term, we can gain a more nuanced understanding of the prevailing sentiment and discussions regarding the Malaysian government.



Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
8
4
6
10
7
1
5
9
2



Topics that appear isolated or separate from the main clusters often symbolize trending or emerging issues. In this particular context, topic 8 stands out as the trending or newly surfaced subject within the collected Twitter posts, specifically representing the Malaysian political landscape.

When we look at the "Top-30 Most Relevant Terms for Topic 8" bar chart, certain terms, such as 'Budget' and 'billion', are noticeably dominant. The prominence of these terms leads us to infer that topic 8 revolves around the economic status of Malaysia.

Therefore, based on this analysis, it's reasonable to conclude that the economic condition of Malaysia is the currently trending or newly surfaced topic of discussion on Twitter. This indicates a strong interest or concern among Twitter users about the economic matters of the country.

Step 4: Sentiment Analysis

Sentiment Analysis of Political Discourse: Sentiment analysis can **determine the overall sentiment (positive, negative, or neutral) expressed** in tweets related to Malaysian politics. This analysis can provide insights into the public sentiment towards political events helping to gauge the overall mood and perception of the political landscape on Twitter.

Approach:
We have developed an approach for sentiment analysis of political events in Malaysia, as outlined below:

Lexicon-based approach using SentiWordNet

This approach utilizes SentiWordNet, a lexical resource that assigns sentiment scores to words based on their semantic relationships. The sentiment analysis process involves mapping words from the input text to their corresponding entries in SentiWordNet and calculating an overall sentiment score. The sentiment score can be positive, negative, or neutral, indicating the polarity of the sentiment expressed.

The approach is relatively straightforward and interpretable. It relies on pre-defined sentiment scores for words, which helps in understanding the sentiment of individual words. However, it may face challenges in accurately capturing the sentiment of complex or context-dependent phrases and expressions.

The sentiment score we obtain ranges from -1 to 1. To classify the sentiment, we divide this range into three classes:

1. -1 to -0.5: **Negative Sentiment**
   The range from -1 to -0.5 is chosen to represent a clear distinction for negative sentiment. By including values below -0.5, it ensures that strongly negative sentiments are captured separately from less negative sentiments. This range allows for a focused identification of tweets expressing strong disapproval, dissatisfaction, or negative emotions.
2. -0.5 to 0.5: **Neutral Sentiment**
   The range of -0.5 to 0.5 as the neutral sentiment class includes considerations for potential non-English or unknown words. This range encompasses values close to zero, which indicates a lack of strong positive or negative sentiment. By categorizing these values as neutral, it accounts for the possibility that tweets within this range may contain words that are not in English or words that we are not familiar with. Thus, the neutral class provides a space to include tweets that do not exhibit discernible positive or negative sentiment due to language variations or unknown terms, ensuring a comprehensive coverage of sentiment analysis while acknowledging potential language challenges or unfamiliar expressions.
3. 0.5 to 1: **Positive Sentiment**
   The range from 0.5 to 1 is chosen to represent positive sentiment. Similar to the negative sentiment range, this range allows for a clear distinction between less positive

sentiments and strongly positive sentiments. By including values above 0.5, it enables the identification of tweets expressing positive emotions, satisfaction, approval, or other positive sentiments.

To calculate the overall sentiment, we exclude the neutral sentiment class, as it may contain words that are not in English or words that we are not familiar with. We then calculate the average sentiment by taking into account the number of tweets in the negative and positive sentiment classes. This is done using the following formula:

$$\bar{x}_{\text{combined}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$\bar{x}_{\text{combined}} = \text{average combined polarity score}$

$n_1 = \text{number of negative tweets}$

$\bar{x}_1 = \text{average polarity score of negative tweets}$

$n_2 = \text{number of positive tweets}$

$\bar{x}_2 = \text{average polarity score of positive tweets}$

Limitations: The lexicon-based approach heavily relies on the availability and accuracy of sentiment scores in SentiWordNet. It may struggle with new or domain-specific words that are not present in the lexicon. Not to mention that it might not capture the nuances of sarcasm or irony since it primarily focuses on word-level sentiment analysis.

Events:

1. #MalaysiaMadani

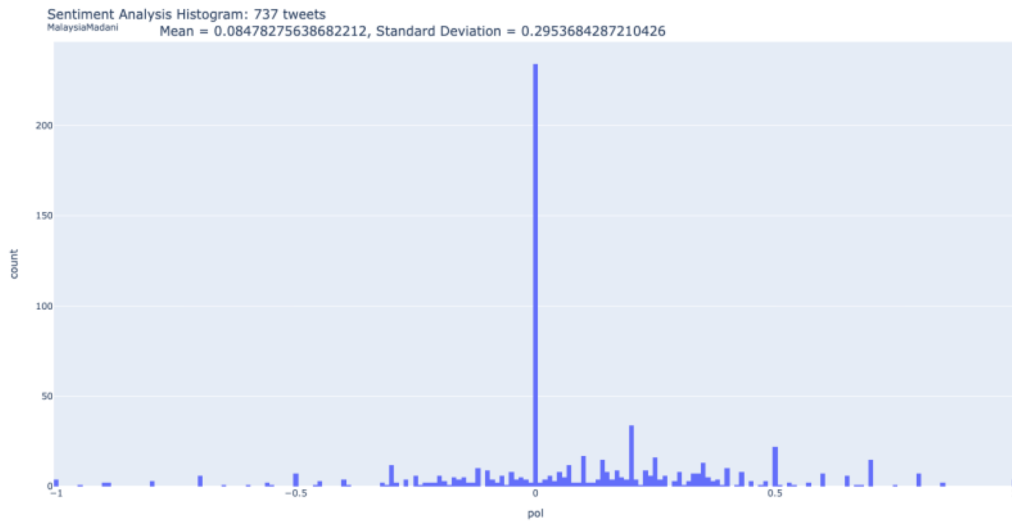


Figure 4.1.1

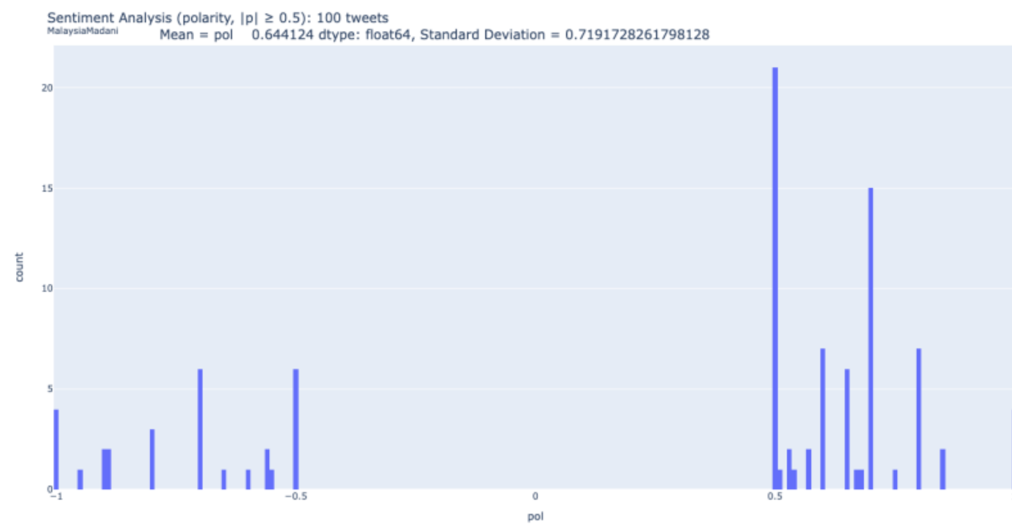


Figure 4.1.2

Overall sentiment : Positive
Average sentiment score : 0.644124

The first event analyzed is #MalaysiaMadani. Malaysia Madani is a government slogan introduced byPrime Minister Anwar Ibrahim on January 19 2023. As each new government faces backlash during the period of government changes between 2020-2023, we decided to choose this event as we believe it to be quite polarizing.

In Figure 4.1.1, this is the overall sentiment analysis of the tweets related to #MalaysiaMadani, however observations suggest that the number is skewed towards neutral due to the amount of words that are unrecognized by the lexicon-based approach. This is most likely due to the tweets being in another language (Bahasa Malaysia).

Figure 4.1.2 excludes the neutral tweets and as we can see, the counts of polarity on the positive side far outweighs the negative side, signaling that the overall sentiment is likely to be positive. This is proven to be true as the mean is 0.644124, which is considered an overall positive sentiment score. It should be noted that the small dataset is due to these being the only 100 available tweets regarding #MalaysiaMadani that had words the lexicon-based approach could identify.

2. #PRU15

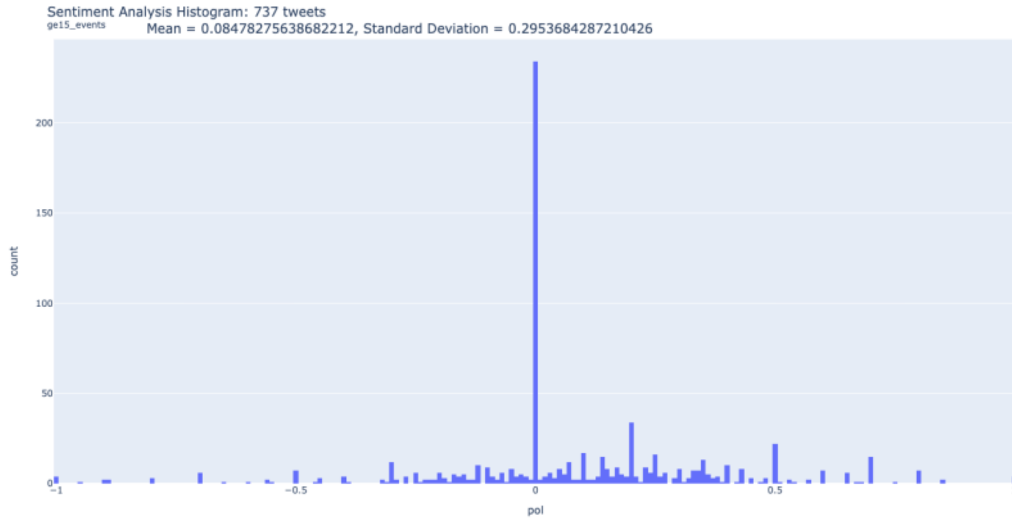


Figure 4.2.1

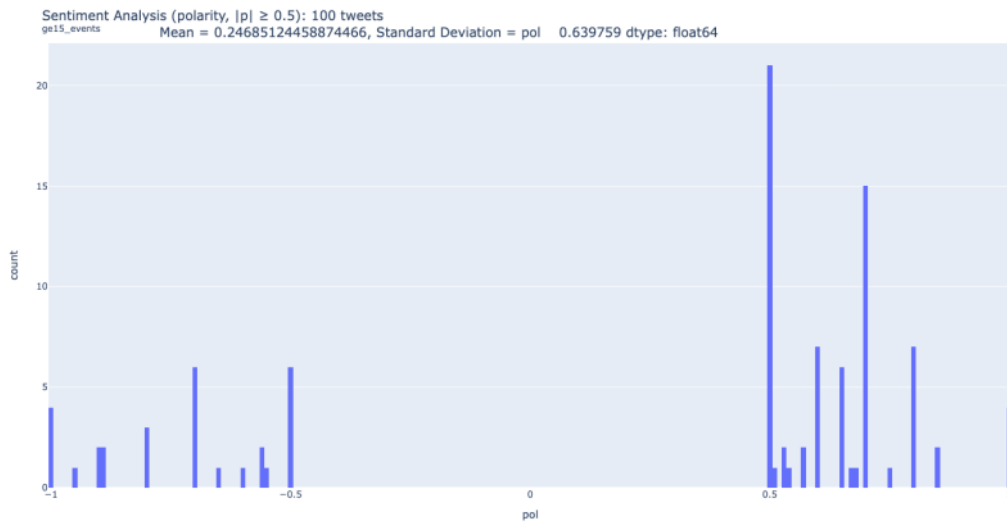


Figure 4.2.2

Overall sentiment : Neutral / Slightly Positive
Average sentiment score : 0.24685

#PRU15 represents all tweets related to GE15 events. General Election 15 was held in November 2022 and represents the chance for our nation to exercise our democracy and vote for its next leader. As each new government faces backlash during the period of government changes between 2020-2023, we decided to choose this event as we believe it to be quite polarizing.

In Figure 4.2.1, this is the overall sentiment analysis of the tweets related to #PRU15, however observations suggest that the number is skewed towards neutral due to the amount of words that are unrecognized by the lexicon-based approach. This is most likely due to the tweets being in another language (Bahasa Malaysia).However, it should be noted that by purely excluding the tweets with 0 sentiment, the distribution shows a slight favor to the positive side of sentiment.

Figure 4.2.2 excludes the neutral tweets and as we can see, the counts of polarity on the positive side slightly outweighs the negative side, signaling that the overall sentiment is likely favoring the positive side. This is arguably true depending on your definition of positive, as the mean is 0.24685, which we consider to be leaning towards the neutral score. It should be noted that the small dataset is due to these being the only 100 available tweets regarding #PRU15 that had words the lexicon-based approach could identify.

3. #KerajaanGagal

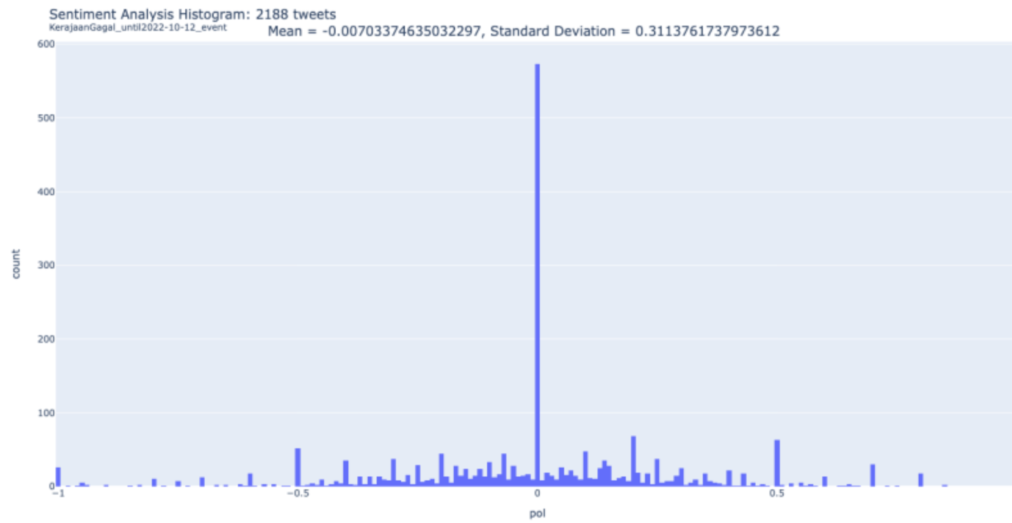


Figure 4.3.1

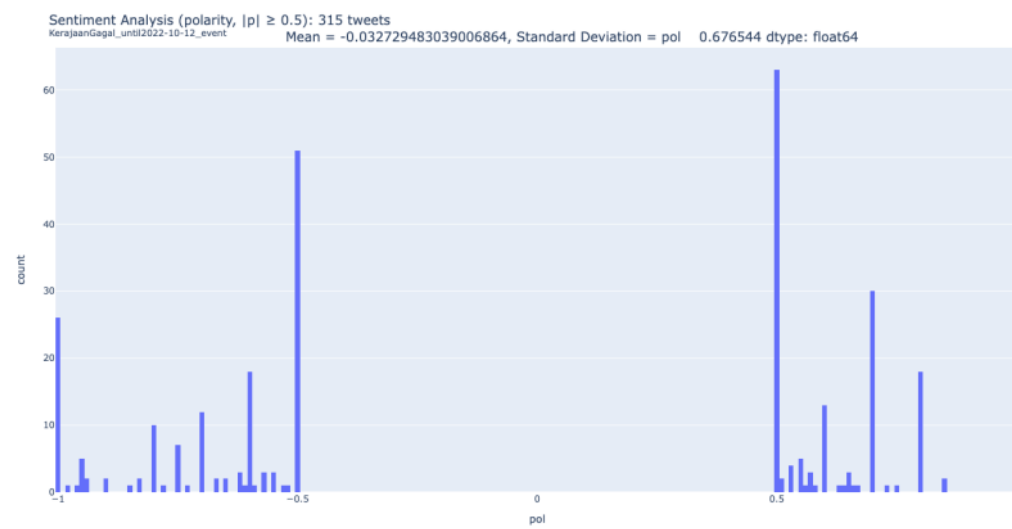


Figure 4.3.2

Overall sentiment : <u>Neutral</u>
Average sentiment score : -0.0327

#KerajaanGagal represents all tweets related to the term coined due to the mistrust in our government after former Prime Minister Muhyiddin Yassin rose to power. As this hashtag was used to show our nation's mistrust towards our own government, we hypothesize the overall sentiment score to be leaning towards the negative side.

In Figure 4.3.1, this is the overall sentiment analysis of the tweets related to #KerajaanGagal, however observations suggest that the number is skewed towards neutral due to the amount of words that are unrecognized by the lexicon-based approach. This is most likely due to the tweets being in another language (Bahasa Malaysia).However, it should be noted that by purely excluding the tweets with 0 sentiment, the distribution does not show any favor towards any side and is likely to be neutral.

Figure 4.3.2 excludes the neutral tweets and as we can see, the counts of polarity on the positive side is eerily similar to the negative side, signaling that the overall sentiment is likely to be neutral. Seeing as the mean is -0.0327, although it is a negative value, it is considered too small to have any effect, which is why we consider this to be neutral.

4. #LangkahSheraton

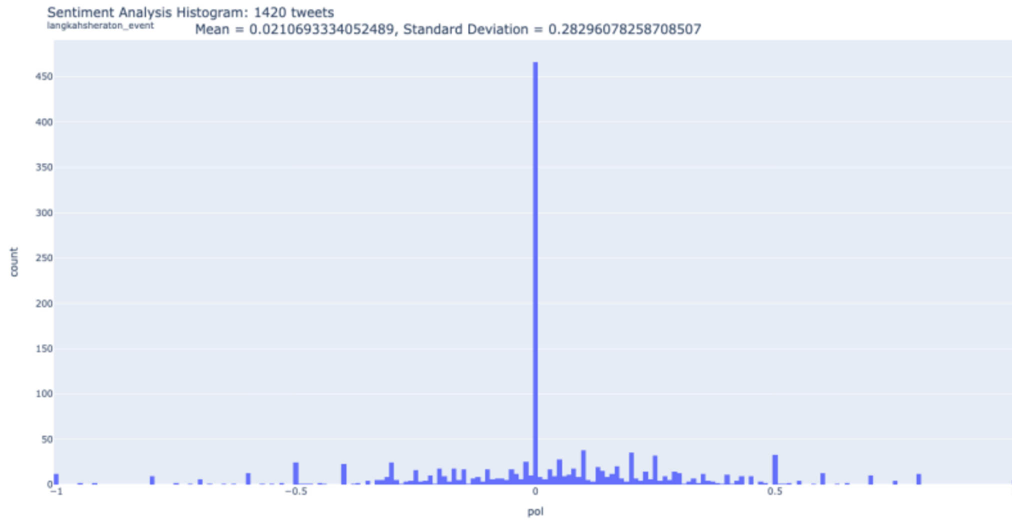


Figure 4.4.1

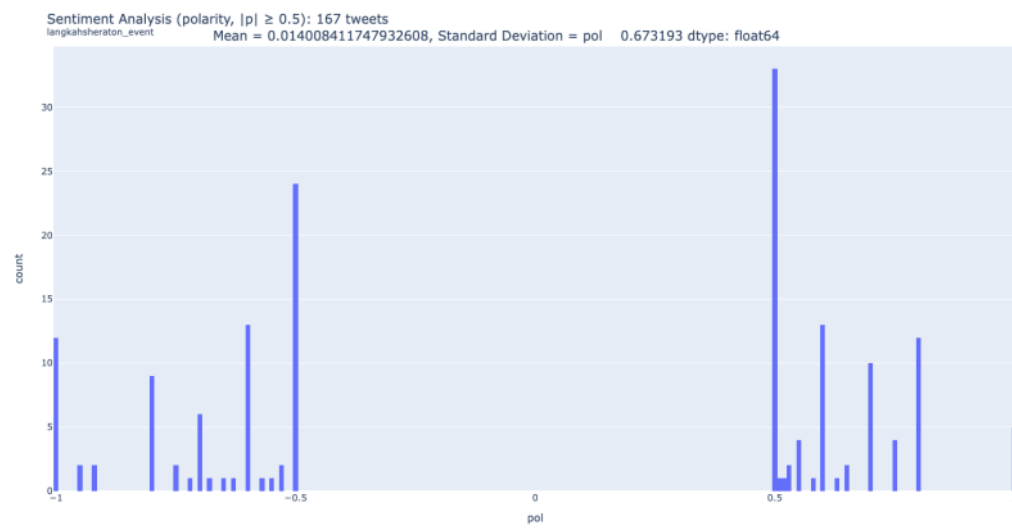


Figure 4.4.2

Overall sentiment : Neutral
Average sentiment score : 0.014008

#LangkahSheraton represents the fall of the ruling Pakatan Harapan (PH) coalition government and the resignation of Prime Minister Mahathir Mohamad after 22 months in power. They were replaced by the Perikatan Nasional (PN) coalition government led by Prime Minister Muhyiddin Yassin. Considered to be one of the largest political events in recent Malaysian history, we chose this event in hopes of being able to scrape a larger number of tweets.

In Figure 4.4.1, this is the overall sentiment analysis of the tweets related to #KerajaanGagal, however observations suggest that the number is skewed towards neutral due to the amount of words that are unrecognized by the lexicon-based approach. This is most likely due to the tweets being in another language (Bahasa Malaysia).However, it should be noted that by purely excluding the tweets with 0 sentiment, the distribution does not show any favor towards any side and is likely to be neutral.

Figure 4.4.2 excludes the neutral tweets and as we can see, the counts of polarity on the positive side is eerily similar to the negative side, signaling that the overall sentiment is likely to be neutral. Seeing as the mean is 0.014008, although it is a positive value, it is considered too small to have any effect, which is why we consider this to be neutral.

5. Najib Razak & #1MDB

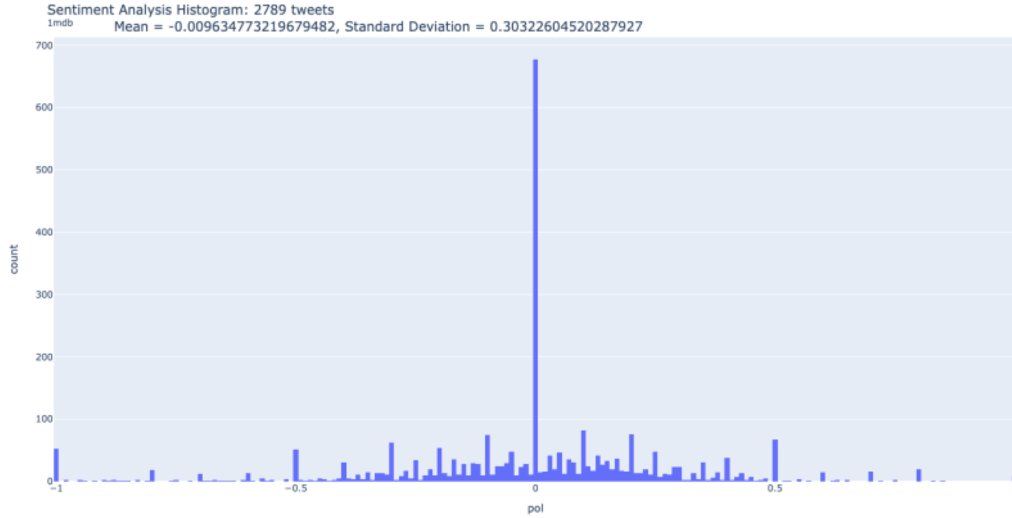


Figure 4.5.1

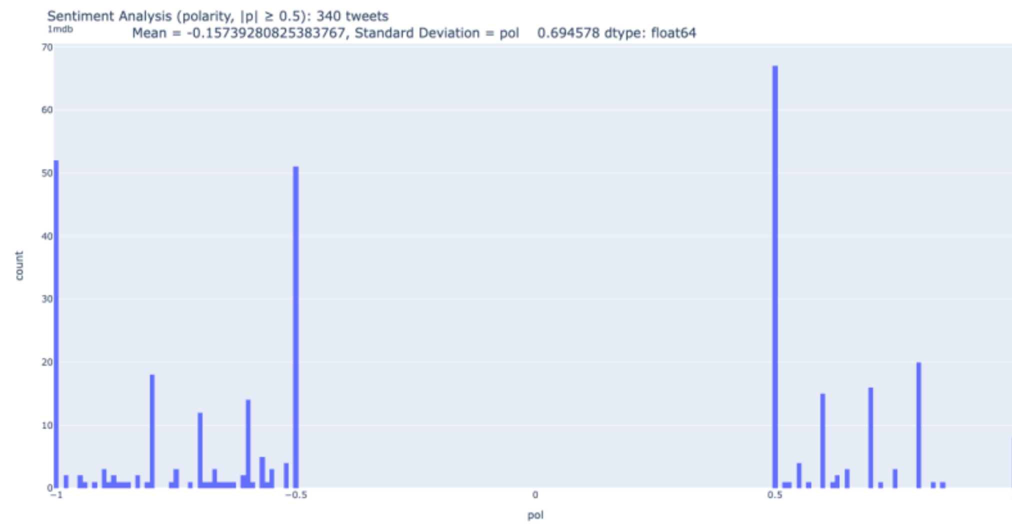


Figure 4.5.2

Overall sentiment : Neutral / Very Slightly Negative
Average sentiment score : -0.15739

#1MDB represents the scheme which involved a systematic embezzlement of funds from the Malaysian sovereign wealth fund 1Malaysia Development Berhad (1MDB) through corruption, bribery, and money laundering. The perpetrators of the scheme diverted the assets globally. Most Malaysians consider this to be the largest political event that has been talked about in the past 20 years, and it is not for good reason. The scandal has rightfully elicited a negative response from the community, to which we have reason to hypothesize that the overall sentiment for this event would be negative.

In Figure 4.5.1, this is the overall sentiment analysis of the tweets related to #1MDB, however observations suggest that the number is skewed towards neutral due to the amount of words that are unrecognized by the lexicon-based approach. This is most likely due to the tweets being in another language (Bahasa Malaysia).However, it should be noted that by purely excluding the tweets with 0 sentiment, the distribution does not show any favor towards any side and is likely to be neutral.

Figure 4.5.2 excludes the neutral tweets and as we can see, the counts of polarity on the positive side is eerily similar to the negative side, signaling that the overall sentiment is likely to be neutral. Seeing as the mean is -0.15739, although it is a negative value, it is considered too small to have any effect, which is why we consider this to be neutral.

An Analysis On The Political Landscape In Malaysia Based On Twitter

ORIGINALITY REPORT

5%	2%	1%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	wikizero.com Internet Source	1%
2	Submitted to M S Ramaiah University of Applied Sciences Student Paper	1%
3	Submitted to Colorado Technical University Student Paper	1%
4	Submitted to Symbiosis International University Student Paper	1%
5	Submitted to The University of the West of Scotland Student Paper	<1%
6	Submitted to University of Malaya Student Paper	<1%
7	en.wikipedia.org Internet Source	<1%
8	www.coursehero.com Internet Source	<1%

9	sciendo.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		