SD2024040000261

ANALISIS TOPIK PERCAKAPAN DAN SENTIMEN PUBLIK PADA *TWEET* TERKAIT PEMILIHAN UMUM PRESIDEN DI INDONESIA TAHUN 2024

Oleh:

M.Alif Al Hakim

Rakha Abid Bangsawan

Tengku Laras Malahayati

SEMIFINAL

BIG DATA CHALLENGE

2024

ANALYSIS OF CONVERSATION TOPICS AND PUBLIC SENTIMENT ON *TWEETS* REGARDING THE PRESIDENTIAL ELECTION IN INDONESIA IN 2024

Abstract

The development of social media marks the dynamics of the communication process, including the role of social media such as Twitter in activities during the presidential election Indonesia in 2024. Politicians, presidential and vice presidential candidates, and the public meet in a communication space to channel aspirations, views and opinions about various issues related to the 2024 presidential election. This research is analysis of conversation topics and public sentiment on ***tweets*** related to the presidential election in Indonesia in 2024 which was carried out using the topic extraction method, sentiment analysis, and social network analysis. Topic extraction is done by using the BERTopic ***framework .*** Meanwhile, sentiment analysis is carried out by using ***fine tuning*** of Transformers models . In addition, social networks were analyzed based on the users who ***were retweeted*** and the topic of conversation. The results of this study shows that the debate between candidate pairs is the most important moment influencing public opinion and responses, and identifying various patterns dissemination of information from each candidate pair. This study concludes the importance of focusing on disseminating work programs and preventing engagement on issues sensitive for each candidate pair.

Keywords: Sentiment Analysis, Social Network Analysis, Topic Extraction, Election President of Indonesia in 2024

I. INTRODUCTION

A. Research Background

In recent years, social media has experienced rapid growth in the number of users and changing the way of distribution information in everyday life (Wigand et al., 2010). Before the existence of the development of social media, the dissemination of information is often hampered by costs the infrastructure needed to reach a wide audience. Currently, with widespread access to the Internet, these barriers have largely been overcome. resolved through the use of social networking sites ***(Social Network Services/ SNS)*** (Kaplan & Haenlein, 2010). In addition, along with the development of media

social, search process, information gathering, and the ability to collecting and sharing opinions and ideas on various topics has changed significantly (Agrawal et al., 2011). As a result, the increase The relevance of communication in social media marks a change in the process public communication that was previously controlled by certain actors, such as politicians, companies, and journalists (Chadwick & Howard, 2009). This phenomenon has become the focus of attention of experts from various disciplines. to understand how digital communication occurs, including in opinion formation on social media.

The rapid use of social media in various aspects of life is visible also in the world of politics, both by the general public and political institutions such as politicians and political parties. Social media becomes a tool and source information that is considered ideal for voicing political aspirations related to policies and political positions of politicians and to build support society towards candidates running for public office (Zeng et al., 2010). Several studies have shown the need for to continuously collect, monitor, analyze, summarize, and visualize relevant political information from social media in order to improve communication between candidates and the general public and voters (Zeng et al., 2010; Kavanaugh et al., 2011; Stieglitz & Dang-Xuan, 2013). For example, it is important to identify the type of user or political leaders' opinions in order to be able to know the discussion trends that are occurring in their circles. In addition, identify issues, trends, and predict Topics that have the potential to increase in the community are also important steps that need to be taken.

One of the social media platforms that is often used for political communication, especially during election campaigns, is Twitter, which now known as X. The use of Twitter in political campaigns began to steal the public's attention when Barack Obama managed to take advantage of it effectively in the 2008 United States presidential election (Wattal et al., 2010). Obama's success is considered an inspiration for politicians in worldwide to consider using Twitter as part of their

their campaign strategy. Meanwhile in Indonesia, Twitter is used in the 2014 and 2019 general elections between presidential candidate Joko Widodo and Prabowo Subianto, as well as in the 2017 Jakarta gubernatorial election between Anies Baswedan and Basuki Tjahaja Purnama.

The use of social media as a means of informing various things, including in the political field, have triggered the emergence of studies regarding social media analytics ***(Social Media Analytics).*** From observations and literature search, it seems that the discussion is about social media analysis still limited. Several studies have conducted social media analysis among them are Akbar et al. (2021), Tumasjan et al. (2010), and Allcott and Gentzkow (2017). Akbar et al.'s (2021) research examines the use of Twitter as a campaign medium in the 2020 United States presidential election, with a focus on analyzing political issues, SARA, as well as themes and sentiments campaign related to the presidential candidate at that time, namely Donald Trump Trump and Joe Biden. Tumasjan et al.'s (2010) study investigated whether Twitter is used as a forum for political considerations and whether the topic The discussion in ***tweets*** on Twitter reflects the political situation during the election federal in Germany in 2009. Allcott and Gentzkow's (2017) research examines the influence of the spread of hoax news during the United States presidential election 2016. From several studies on social media analysis, there are research gaps in sentiment analysis and ***social networks*** in use of social media Twitter where all three show the use in an overseas context. Study of social media analysis in the world politics in Indonesia is still limited, even when viewed from the number of people users, Indonesia is classified as a country with a large number of social media users. the world's largest social media (We Are Social, 2024). It is in this position that this research carried out by focusing on the presidential election campaign platform in Indonesia in the 2024 presidential election.

When Indonesia holds presidential elections in 2024, Twitter become a platform that is often used by Indonesian people to give opinions about presidential and vice presidential candidate pairs. Indonesian people are actively expressing their views on issues

relevant politics, as well as to review the performance and programs of each presidential candidate pairs. Voter behavior, communication between candidate pairs the president and vice president with prospective voters, as well as politicians in the 2024 Indonesian Presidential Election which was revealed in the media Twitter social media has become something that enlivens the use of social media in Indonesia.

Utilization of topic extraction and sentiment analysis technology Using ***Artificial Intelligence*** can help the campaign team or the candidate pair themselves in determining a suitable strategy in facing the 2024 Presidential Election. By developing a framework work to analyze topics and sentiments for election purposes, parties related to the need can save time and costs in the process strategy or decision making.

B. Problem Formulation

Based on the background that has been explained, the problem formulation is The research team will propose the following.

1. What topics of conversation do Twitter users frequently talk about? regarding the 2024 Indonesian Presidential Election?
2. What kind of moments or events get a lot of response from society during the 2024 Indonesian Presidential Election?
3. What kind of conversation topics tend to get a good response? positive from the community during the 2024 Indonesian Presidential Election?
4. Does the number ***of followers*** have a big influence on the distribution of information on a topic?
5. Are there any users who are key actors in the distribution? information on a topic?

C. Research Objectives

This research was conducted to achieve several objectives as follows.

1. Know the topics of conversation that users often talk about Twitter regarding the 2024 Indonesian Presidential Election.
2. Know what moments get a lot of response from the public. during the 2024 Indonesian Presidential Election.

3. Know the types of conversation topics that tend to get positive response from the public during the Indonesian Presidential Election this year 2024.
4. Knowing the influence of the number ***of followers*** on the spread of information a topic.
5. Knowing whether there are users who are key actors in dissemination of information on a topic.

D. Benefits of Research

The research conducted has the following benefits.

1. Support the development of Computer Science in understanding various areas of community life including socio-political areas. Thus, Computer Science is not only examined from the side application development alone, but also needs to be developed for support human life in general.
2. Opening up the potential for developing studies on ***Social Media Analytics,*** especially for the presidential election.
3. Encourage the development of analytical thinking and studies on use of social media in presidential election campaigns.
4. Be a source of aspiration for politicians in preparing their campaigns. politics by utilizing ***Social Media Analytics.***

II. METHODOLOGY

A. Experimental Method

This study uses a ***dataset*** containing a collection of Indonesian people's interaction on social media Twitter, which is now known under the name X, during the 2024 Indonesian Presidential Election campaign. The ***dataset*** contains ***tweets*** related to three candidate pairs. president and vice president in the 2024 Indonesian Presidential Election. Third The couple is:

- Candidate pair number 01: Anies Baswedan and Muhaimin Alexander
- Candidate pair number 02: Prabowo Subianto and Gibran Raka Abuming
- Candidate pair number 03: Ganjar Pranowo and Mahfud MD

The following is an overview of the research flow carried out by the research team.

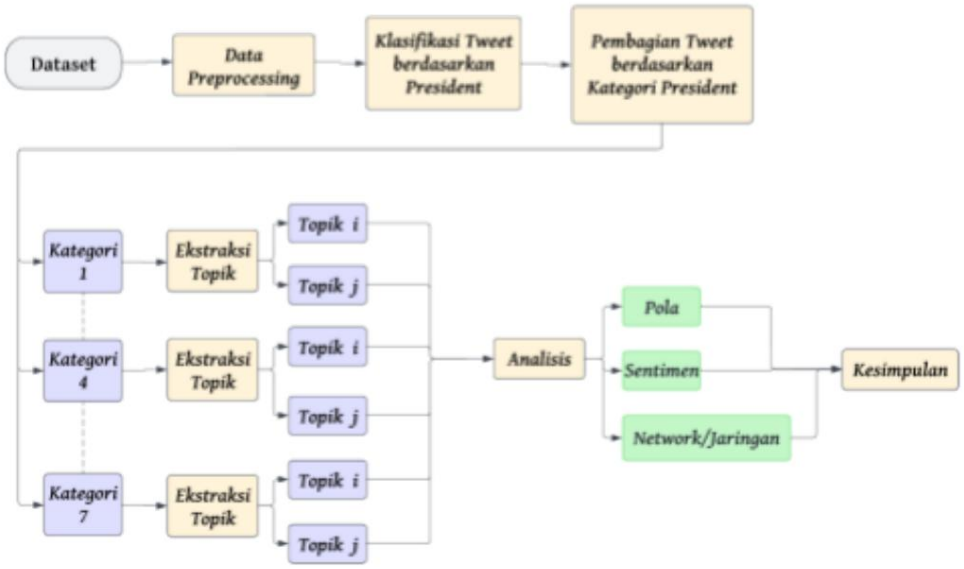


Figure 1: Research Flow

In this study, a pattern analysis will be carried out on topics, sentiments, and the structure of the interaction network found in each topic of discussion. The purpose This analysis is to dig up information on hot topics. discussed, identifying the reactions of the majority of Indonesian society towards the topic, and identify users who play an important role in dissemination of information related to a topic.

In general, research begins by conducting data preprocessing. on the initial ***dataset .*** The data preprocessing stage includes the data cleaning process ***tweet*** text , column deletion, and handling of ***tweet*** data that has ***missing values.*** Each ***tweet*** is then classified based on the pair candidates under discussion. The classification process involves modeling that use key words related to each pair

candidates. The classification results obtained group each ***tweet*** into seven categories based on the candidate pairs discussed in the ***tweet .*** The following is an overview of the division of these categories.

	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7
Partner Candidate 1	ÿ			ÿ	ÿ		ÿ
Partner Candidate 2		ÿ		ÿ		ÿ	ÿ
Partner Candidate 3			ÿ		ÿ	ÿ	ÿ

Table 1: Dataset Category Distribution

This category division is done to help find patterns that its nature is more detailed and avoids findings that are global or too general.

Each category is then subjected to topic extraction to find patterns of conversation in society when discussing certain candidate pairs. The topic extraction process is carried out using the BERTopic.1 ***framework .*** Next, each topic will be analyzed for its patterns and relationships with features. others in ***the dataset,*** such as ***follower*** features , number of ***retweets, timestamps,*** and others. In addition, social network analysis and sentiment analysis will also be carried out. on each topic to discover additional insights.

B. Dataset

1. ***Dataset*** Description

The dataset used in this research is a collection interactions from social media Twitter related to three presidential candidate pairs during the 2024 Indonesian Presidential Election campaign, which provided by Media Kernels Indonesia.2 This data was obtained via ***scraping*** techniques on social media Twitter. The results of this ***scraping*** produces a large ***dataset*** consisting of several columns, including:

[1] https://maartengr.github.io/BERTopic/
[2] Dataset Description on the Satria Data ***Website*** Page

Column Name	Column Description
created_at	time ***the tweet*** was created
username	encrypted Twitter ***username***
content	***tweet*** text
interaction_type	types of user interactions ***(retweets, mentions,*** or ***reply)***
following_count	number ***of following*** (friends)
followers_count	number of ***followers***
tweet_count	the number of ***tweets*** that have been created by user
location	location listed in profile user
language	language detected from ***tweet*** content

Table 2: Dataset Description

2. Data Construction

Data Understanding. In this study, the research team was given ***a dataset*** with a total of 9.8 million or 9,817,355 lines of ***tweet*** content related to the three presidential candidate pair during the 2024 presidential election campaign. ***The dataset*** to be analyzed has 12 columns as already mentioned. explained in the Dataset Description section.

The Twitter ***username*** column could not be analyzed by the research team furthermore because the ***username*** has been encrypted and the research team no information is provided regarding the encryption algorithm or ***key*** so that The research team was unable to perform decryption to find out ***the username.*** Twitter that does the interaction.[3]

The research team decided to remove the “type” column, “lang”, and “loc” from ***the dataset*** for some reason. The “type” column has only one value, namely "tweet" so it does not provide

[3] Is it theoretically possible to decrypt a file without the key/password? - Information Security Stack Exchange

data variation and does not have ***missing values*** because the numbers are the same with the number of rows of data (9,817,355). The “lang” column shows 13 different languages according to ISO 639-14 standard , but further analysis shows a lot of language detection errors so that the data is not accurate as in Figure 2. Finally, the loc column has many missing value, about 60% of the total data, and a lot of noise data. generated from users ***customizing*** their location information so it does not provide useful information.

Figure 3 is a visualization of the distribution of data from the number of ***following*** and ***followers*** owned by the user, as well as ***tweets*** that created by users.

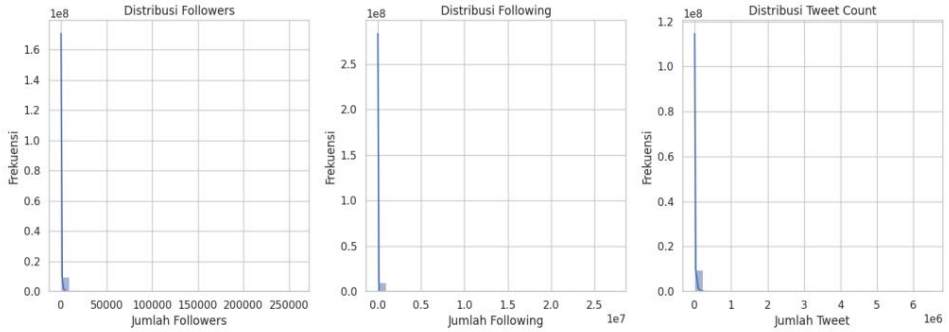


Figure 3: Distribution of ***Following, Followers,*** and Number of User ***Tweets*** Data

Based on Figure 3 and after further analysis, the research team get information that the given dataset contains various user types, starting from the famous (seen by the number of ***followers)*** to the point of not being famous. The research team suspects that users who Having a lot of ***followers*** will have an impact on a topic. being discussed. An example of the impact in question is being able to influence or lead opinions related to the topic being discussed.5

Data Cleaning. Before carrying out further processing stages, The given dataset needs to be cleaned first. Process This preprocessing involves several steps as follows.

[4] ISO 639-1 - Wikipedia
[5] https://www.mdpi.com/2076-0760/12/7/402

1) **Correct or normalize the text in *the tweet.*** There is text data that has problems such as Unicode characters which is invalid. To fix the issue, the team Researchers use the ftfy library (fixes text for you) so that the resulting text is more consistent and interpretable properly by the algorithm or the next stage.

2) **Handling lost data.** Lost data will be deleted. from the dataset.

3) **Handling *noise data. Noise*** data such as in the loc column will removed from the dataset.

4) **Using Regex method and *method* from python.** Method best to make text data in the content column more Consistency is with the Regex method and the python ***method*** such as lower() and replace() which can be seen in Table 3.

One example of the results of the pre-processing process above can be seen in Table 4.

C. Modeling

1. Presidential Classification Model

To classify presidential and vice presidential candidate pairs the president mentioned or discussed in a ***tweet,*** is built a phrase matching algorithm model. This model is designed to recognize various variations of the terms used for candidate pairs, including full name, nickname, abbreviation and name combinations couples. Model development begins with the construction of a dictionary which associates each candidate pair with a list of keywords. related.

The text classification process involves two main steps. First, The model tries to match phrases exactly to a list of words. key. If not found, the model switches to matching with fuzzy algorithm . ***Fuzzy*** matching can help find ***strings*** that are similar but not identical. The matching technique is

The fuzzy algorithm used in this study is the partial ratio algorithm. with a similarity threshold of 80 percent. Then, the indicator binary is created for each candidate pair in order to show whether a ***tweet*** mentions a particular candidate pair or not. This indicator is then used to divide ***the dataset*** into seven different categories as in Table 1.

2. Topic Extraction

Topics are extracted using ***the BERTopic framework.*** However, Before that, ***the dataset*** for each category was performed preprocessing in the form of deleting duplicate ***tweets .*** This is necessary is done so that the model is not biased towards repeated data and prevent the model from failing to generalize a topic. Then, each ***tweets*** are subjected to feature extraction using the model ***Transformers*** IndoBERTweet. This model was chosen because it has been trained on Twitter data in Indonesian so that it has a dictionary that more appropriate and larger for the text data to be analyzed. Results ***embeddings*** then the ***cosine similarity*** calculation is performed between each ***tweet*** to find similar ***tweets*** between one ***tweet*** and another. other ***tweets ..*** This needs to be done because there are ***tweets*** that only differ by one or two words, either because it was deliberately created by the user or due to ***username*** encryption on ***mentions.*** For example, ***tweets*** "@QOS7XYPB Candidate pair 01 has good policies" and "@Mwwr2Fp Candidate pair 01 has good policies" are similar ***tweets*** after encryption even though both are comes from a different Twitter account.

The ***embedding*** results are then subjected to ***dimensionality reduction.*** using UMAP (Uniform Manifold Approximation and Projection). This needs to be done to avoid ***high dimensionality curse.*** The features that have been reduced then grouped into several clusters using the model HDBSCAN. This model is used because it does not impose

each data to be included in a topic category, so that produces categories that contain ***outliers.*** This ensures purity of topics extracted from other ***tweets*** that are not actually related.

The clusters that have been generated are then represented by using ***count vectorizer*** and ***c-TF-IDF.*** In ***the count vectorizer,*** added a list of ***stop*** words so that these words are not are taken into account as a representation of the topic. ***Stopwords*** consist of ***stopwords*** in Indonesian, ***slang,*** and common words related to the 2024 Indonesian Presidential Election, such as name of candidate pair, 'presidential candidate', 'vice presidential candidate', and others. Each keyword then the weight is calculated using ***c-TF-IDF.***

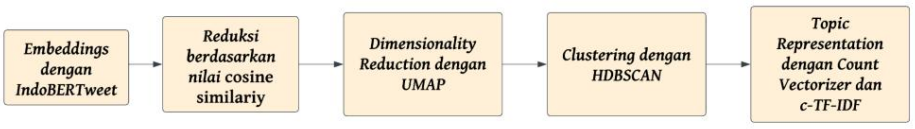


Figure 4: Topic Extraction Flow

The model is then used to predict topics. The entire ***dataset.*** The topic prediction results are then analyzed for their patterns. based on tweet creation time , user followers, number of ***retweets,*** sentiment, and others.

3. Sentiment Analysis

Sentiment analysis was performed using three models different ***pretrained models ,*** namely Indo-RoBERTa,[6] Indo-BERT,[7] And Indo-BERTweet.[8] These three models have been trained for classification tasks. sentiment in Indonesian and shows good performance in previous experiments based on the results that have been obtained

[6] w11wo/indonesian-roberta-base-sentiment-classifier · Hugging Face
[7] mdhugol/indonesia-bert-sentiment-classification · Hugging Face
[8] ridhodaffasyah/sentiment-analysis-indobertweet · Hugging Face

conducted on the Hugging Face platform. Each topic will be classified sentiment based on each ***tweet*** it uses these models.

The use of several of these models aims to achieve triangulation of results. If the three models give similar results, this is can increase confidence in the validity of sentiment predictions that was done. In addition, a comparison between the prediction results of the three models can provide additional insight into the advantages and disadvantages of shortcomings of each model in the context of tweet ***data*** analyzed.

4. Social Network Analysis by Uploader Account

In the dataset there is a ***username*** column and ***a mention column*** in the column ***content.*** However, the ***username*** has been encrypted so it is not allows for social network analysis. Although Thus, each ***retweet*** has a code at the end of ***the tweet*** as follows "[re uploader_account_name]". The research team utilized the information This is to perform analysis on the uploader's account regarding the topic. which has been extracted.

The purpose of this analysis is to identify accounts which plays a role in the dissemination of information on a topic. It is necessary remember that information spreads the most through ***retweets*** so that the uploader accounts can become key actors in the distribution of certain topics. Thus, this analysis can helps understand the important role of certain accounts in the dynamics dissemination of information regarding the 2024 Indonesian Presidential Election on social media Twitter.

III. RESULTS AND DISCUSSION

Based on analysis of topics, sentiment, and social networks has been carried out, the research team obtained several results and findings that interesting. These results and findings are also associated with supporting data in the form of

news and information circulating in the news about each couple candidates during the 2024 Indonesian Presidential Election campaign period.

1. Presidential and Vice Presidential Candidate Debate

During the campaign period for the 2024 Indonesian Presidential Election, the Commission The General Elections Commission (KPU) held a debate between presidential candidate pairs and vice president. This event serves as a platform to disseminate profiles, visions, missions and programs of candidate pairs to voters and society (Wahyuni, 2024). In addition, this debate provides information to voters as one of the considerations in determining choices, as well as delving deeper and elaborating on each theme. raised in the election campaign.

This debate event is also a means for the public to find out views of candidate pairs regarding issues concerning national life and in Indonesia. This year, the candidate debate event was held as many as five times with several different themes, covering law and Human Rights (HAM), defense and security, economy and trade, energy and natural resources (SDA), and information technology and employment (Kulon Progo Kesbangpol Agency, 2023).

This debate became one of the events that invited reactions. society. This can also be seen from the results of ***the dataset analysis.*** Based on analysis of ***dataset*** categories 4, 5, 6, and 7, many Twitter users who discuss and form opinions about the three candidate pairs. From In the discussion, the research team found that the debate between the couple Candidate numbers one and two are topics of discussion that are often discussed people on Twitter.

In addition, in figure 5, there is a lot of talk about state defense equipment and land percentage. Both topics are themes the third and fourth candidate debates discussed defense and security and energy and natural resources. National defense data is also one of the topic of conversation among Indonesian people on Twitter. The topic became

one of the hot topics at that time because there was a discourse about openness of information regarding data and facts on national defense and security (Kompas.com, 2024). There was a spike in the number of ***tweets*** around the 7th and 8th. January 2024 also serves as supporting evidence that candidate debates are at the center of attention. public attention due to the debate on national defense and security held on January 7, 2024.

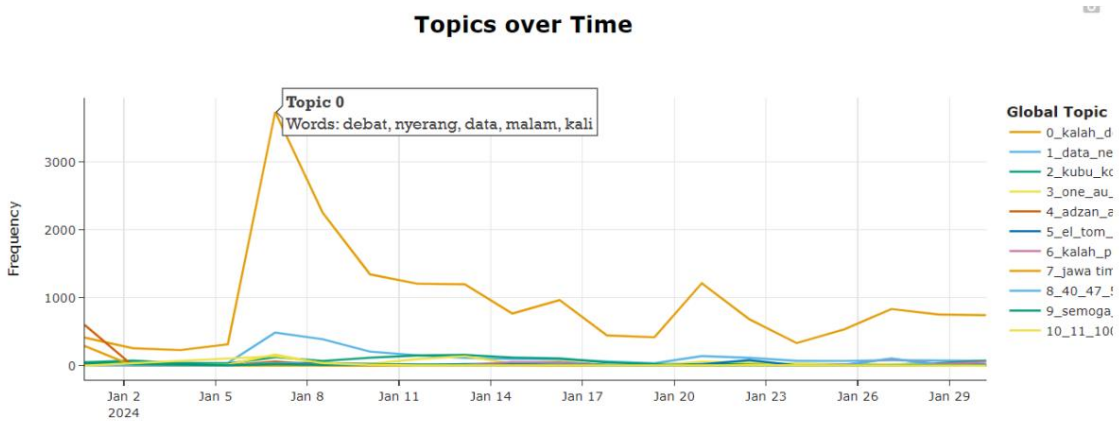


Figure 6: Development and Frequency of Topics Over Time for Category 4 Dataset

Based on the topics discussed in Figure 5, it can be seen that The majority of ***tweets*** discussed candidate pair number 2. Results sentiment classification also shows that most of the topics are have negative sentiment results.

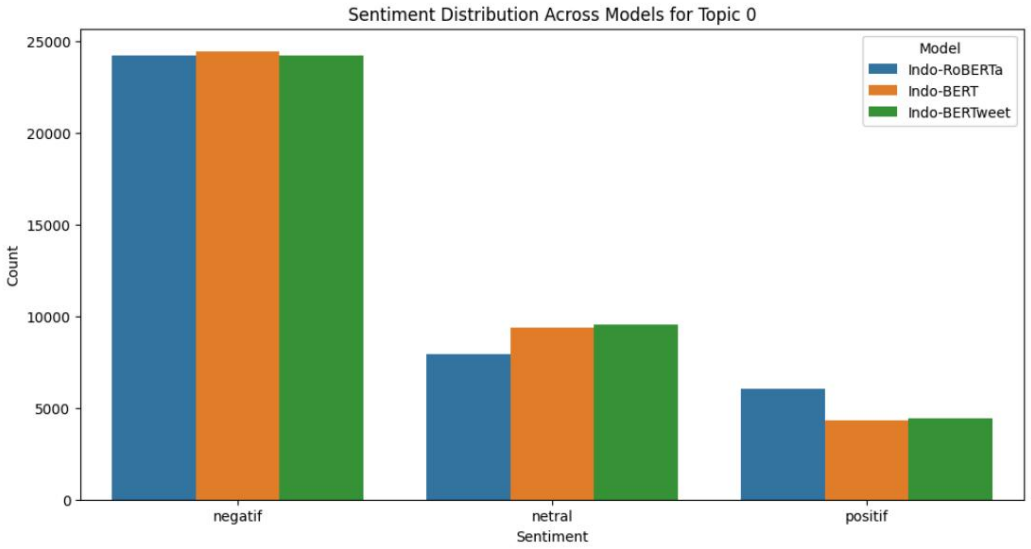


Figure 7: Topic Sentiment Classification 0 Dataset Category 4

From the description and visualization above, the research team concluded that the debate presidential and vice presidential candidates influence Indonesian public opinion on social media Twitter.

2. Campaign Style for Each Candidate Pair

Campaign method or style regarding candidate pair number one become one of the topics that are often discussed. In figure 8, it can be seen that The word 'desak' plays an important role in forming topic 0 regarding the campaign candidate pair number one. This refers to the campaign style candidate pair number one who held a program called Urge Anies. In addition, it is also seen that the candidate pair number one can attract attention from groups of Korean music fans or those who are used to it referred to as ***kpopers.*** However, based on pictures 9 and 10, it appears that there are still many differences of opinion or polarization towards the campaign style. Based on ***the tweets*** in the category 5 ***dataset ,*** This candidate pair received negative sentiment related to their comments towards Rohingya refugees and there are indications of the use of the call to prayer as a campaign tools, as seen in figure 11 and figure 8, as well as the results sentiment analysis in figure 12 and figure 13.

Meanwhile, topic information regarding the candidate pair's campaign number two is not often discussed. However, the candidate pair serial number two often gets attention related to ***tweets*** that discussing humanitarian activities and also about Islamic boarding school students such as seen in figure 14. The comments of this candidate pair are also often became the center of public discussion. The comments included issued during the debate and resulted in negative sentiment from the public.

Candidate pair number three has ***tweets*** that focus on the work programs it has. These programs are centered on to the people, such as farmers and fishermen, and also discusses digital security. Topics regarding this work program have received sentiment good from the community as seen in pictures 15, 16, 17, 18, and 19.

3. The Influence of the Number *of Followers* on the Popularity of a Topic

In this study, the research team realized that the ***retweet*** method become the most widely used method in disseminating information, as in figure 20. However, after conducting the analysis, the research team found that there was no relationship between the number ***of followers*** and popularity of the topic. Users with many ***followers*** do not guarantee ***His tweet*** will ***be retweeted*** by many other users or on the same topic with ***his tweets*** also being discussed by others. Here is one of them visualization that describes the situation.

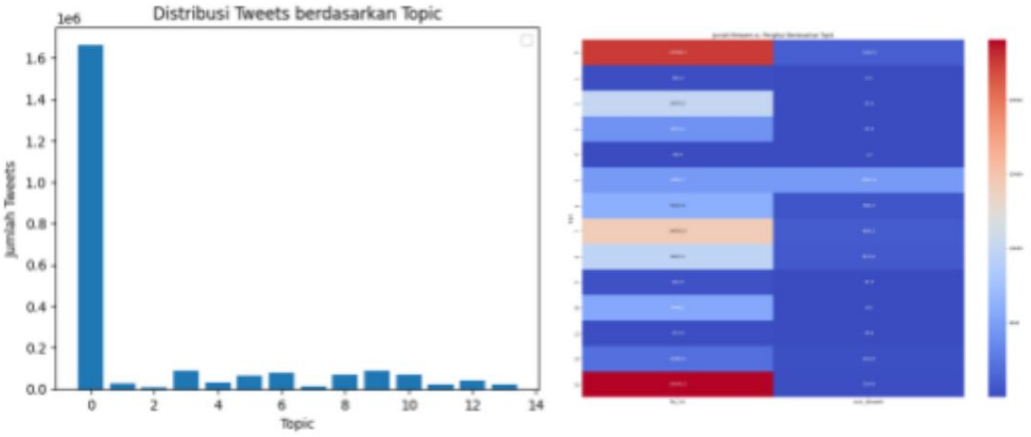


Figure 21: Visualization of Heatmap ***of Followers*** and ***Number of Retweets*** and Topic Frequency

It can be seen that topic 13 is rarely discussed even though the average ***follower*** who owned by users is classified as high.

4. Interaction Network Analysis

Based on the interaction between the uploader's account and the topic discussed, the research team found several patterns as follows.

1. The accounts “mdy_asmara1701” and “aniesbubble” have an important role in disseminating information about the candidate pair's campaign serial number one, as seen in figure 22 interaction network category 1 in the appendix.
2. The “pdi_perjuangan” account has an important role in spreading information regarding the program of candidate pair number three. This is

seen in figure 23, the category 3 interaction network is found in attachment.

3. News accounts, such as “tribunnews”, play an important role in disseminate information on factual issues related to candidate pairs. For example, the topic of land percentage in the data category 4 or issues regarding coalitions between candidate pairs in the data category 5. This can be seen in figure 24 of the category 4 interaction network. and 25 category 5 images in the attachment.

IV. CLOSING

1. Conclusion

Based on the analysis that has been carried out, the research team found several key findings. First, the presidential and vice presidential debates The president became a moment that triggered a big response from users Twitter in Indonesia. Second, the majority of topics discussed are regarding support for the presidential and vice presidential candidate pairs certain presidents. In addition, there is a discussion about the campaign style each pair of candidates, where candidate pair number one highlighting a more contemporary campaign style, candidate pair number sequence two focuses on humanitarian actions, and candidate pairs serial number three emphasizes his work program. Third, no found a significant correlation between the number ***of followers*** and popularity of the topic of conversation. Fourth, key actors often appear in various topics related to candidate pairs or certain issues. Finally, the topic of discussion is oriented towards the interests of the people. tend to have positive sentiment, while related topics with sensitive issues such as religion or group tend to have mixed sentiments.

2. Suggestions

Several suggestions can be generated from the results of the analysis that has been carried out. done. The following are suggestions that the research team can provide.

1. Other researchers can develop similar studies by using Social Media Analytics to enrich the study about social media in various areas of life.
2. In the political field, politicians and interested people running for regional head and presidential elections, or other important positions in government should avoid any comments on sensitive issues because will have an impact on decreasing public perception.
3. Each candidate pair or campaign team should focusing on the dissemination of information about programs to be executed.
4. Candidate pairs should prepare themselves better in facing moments like debates because it will become the center of public attention.
5. The number ***of followers*** of a user does not guarantee increase the popularity of a topic. Therefore, the team The campaign should focus on delivering the program which are interesting.
6. It is important for each candidate pair and campaign team to maintain ***his image*** in front of the press. This is because factual information will spread quickly through those news or press accounts.

V. REFERENCES

Wigand, R.T., Wood, J.D., & Mande, D.M. (2010, 8 5). Taming the Social Network Jungle: From Web 2.0 to Social Media. ***AMCIS 2010 Proceedings, 416.*** https://aisel.aisnet.org/amcis2010/416

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. ***Business horizons,*** 53(1), 59-68. https://www.sciencedirect.com/science/article/pii/S0007681309001232

Agrawal, D., Budak, C., & Abbadi, A.E. (2011). Information diffusion in social networks: observing and influencing societal interests. ***Proceedings of the VLDB Endowments,*** 4(12), 1512-1513.

Chadwick, A., & Howard, P. N. (Eds.). (2009). ***Routledge Handbook of the Internet Politics.*** Routledge.

Zeng, D., Chen, H., Lusch, R., & Li, S.-H. (2010). Social Media Analytics and Intelligence. ***IEEE Intelligent Systems,*** 25(6), 13-16.

Wattal, S., Schuff, D., Mandviwalla, M., & Williams, C. B. (2010). Web 2.0 and Politics: The 2008 US Presidential Election and an E-Politics Research Agenda. ***MIS Quarterly,*** 34(4), 669-688. http://www.jstor.org/stable/25750700

Kavanaugh, A., Fox, E.A., Sheetz, S., Yang, S., Li, L.T., Whalen, T., Shoemaker, D., Natsev, P., & Xie, L. (2011). Social media use by government: From the routine to the critical. In ***Proceedings of the 12th annual international digital government research conference: digital government innovation in challenging times*** (pp. 121-130).

Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. ***Social network analysis and mining,*** 3, 1277-1291.

Akbar, P., Irawan, B., Taufik, M., Nurmandi, A., & Suswanta. (2021). Social Media in Politics: Political Campaign on United States Election 2020 Between Donald Trump and Joe Biden. In ***HCI International 2021 - Late Breaking Posters*** (pp. 359-367). Springer International Publishing.

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In ***Proceedings of the international AAAI conference on web and social media*** (Vol. 4, pp. 178-185).

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. ***Journal of economic perspectives,*** 31(2), 211-236.

Digital 2024: 5 Billion Social Media Users. (2024, 1 31). We Are Social. Retrieved 7 13, 2024, from https://wearesocial.com/id/blog/2024/01/digital-2024-5-billion-social-media-users/

Kompas.com. (2024, January 9). Third Presidential Debate: Defense and Security Data State Secrets. Kompas. https://nasional.kompas.com/read/2024/01/09/07000011/debat-pilpres-ketiga--data-defense-and-security-state-secrets

Wahyuni, W. (2024, January 2). Benefits of presidential and vice presidential debates for prospective voters. Lawonline. https://www. Hukumonline.com/berita/a/ benefit-debat-capres-dan-cawapres-bagi-potential-voters-lt6593e7553434b/

Kulon Progo Kesbangpol Agency. (2023, December 4). Presidential and vice presidential debate theme 2024 Election. Kulon Progo Kesbangpol.

https://kesbangpol.kulonprogokab.go.id/detil/881/tema-debat-capres-cawapres-p

2024-Election

VI. APPENDIX

Regex and Methods Python	Description
r'@[^]+'	Remove mentions from a ***tweet.***
r'http\S+|www\S+'	Removes the URL from ***the tweet,*** even if it starts with . with 'http' or 'www'.
r'\s+'	Replaces multiple spaces with a single space and removes spaces at the beginning and end of text.
r'^rt'	Remove the 'RT' prefix from a ***tweet.***
.replace('\n',' ')	Replaces ***whitespace*** characters ('\n') with spaces (' ').
.lower()	Change all letters in ***a tweet*** to lowercase.

Table 3: Use of Regex and Python ***Methods*** to Extract ***Tweet*** Data

Initial Text	Text After *Preprocessing*
@HUo2P8p6enbSrHhNtJ hdkhUGz+KnIaG9j7IK6 F0uJSg= Sorry yeah.. sorryyy yeah.. Prabowo Gibran never	sorry yeee.. sorryy yeee.. prabowo gibran never...
RT Foreign Media Highlight Anies' Way Attract Young Voters Use 'Urge Anies' https://t.co/UYEMKs	Foreign media highlights Anies' method of attracting young voters using 'pushing Anies'

4Q10	
@5Tz753nhSrc3MyKUjK PGAKQ59HMT5TnaXVJrY BVFXro= Team AniEs also has tofa lemon, no less its racist	Anies' campaign team also has Tofa Lemon, which is no less racist

Table 4: Example Results of Data Preprocessing

Figure 2: Example of Language Detection Error Sample

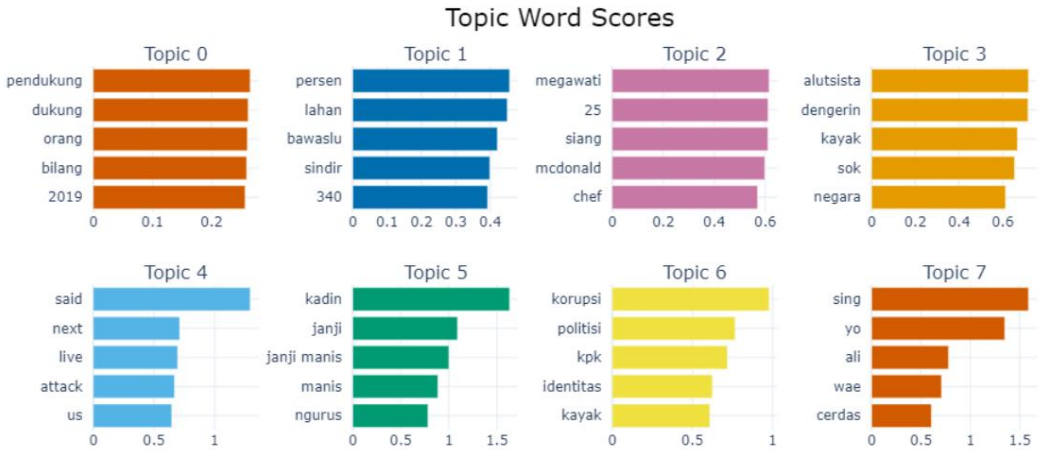


Figure 5: List of Discussion Topics Regarding Category 4 Data

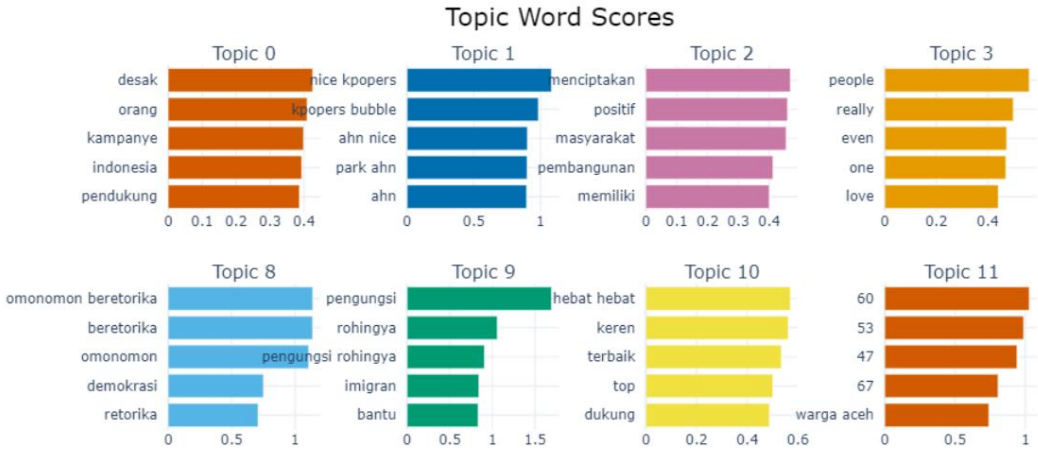


Figure 8: List of Discussion Topics Regarding Category 1 Data

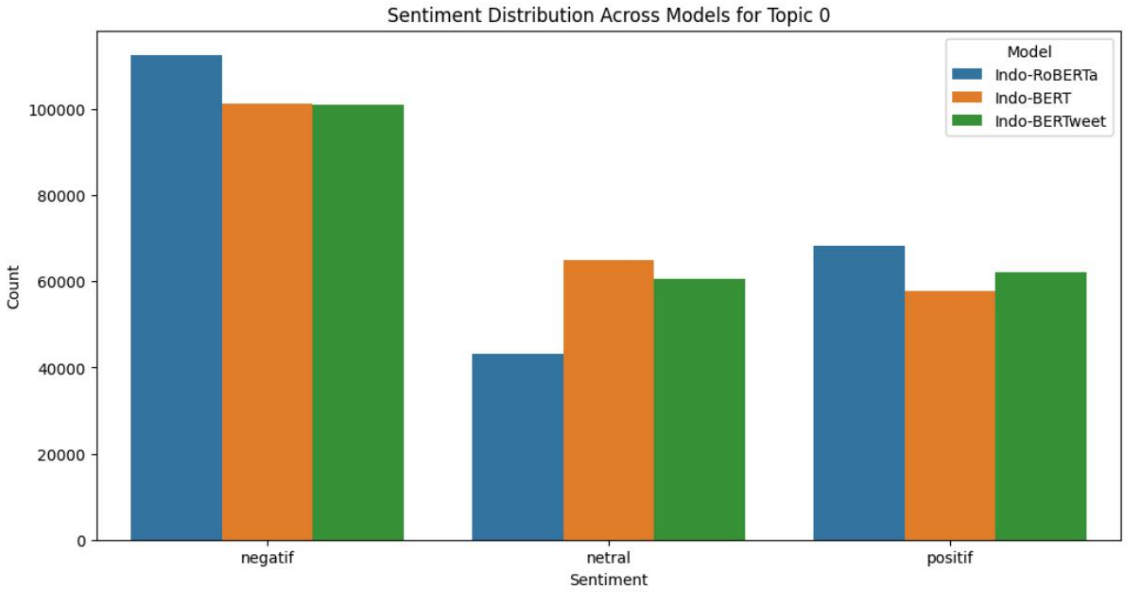


Figure 9: Distribution of Sentiment Topics Regarding Category 1 Data About Topic 0

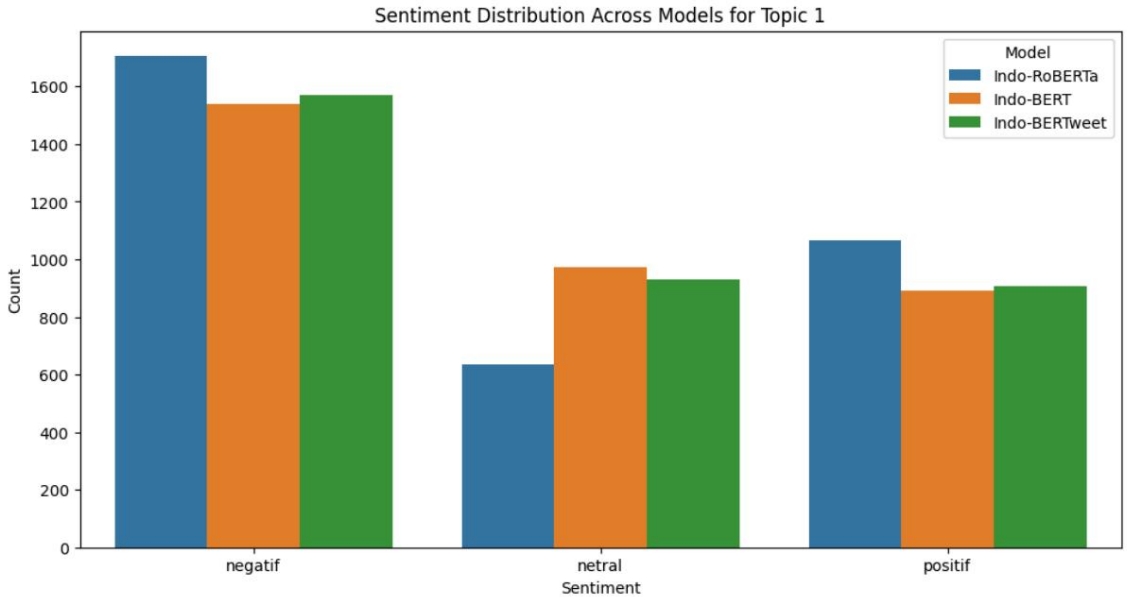


Figure 10: Distribution of Sentiment Topics Regarding Category 1 Data About Topic 1

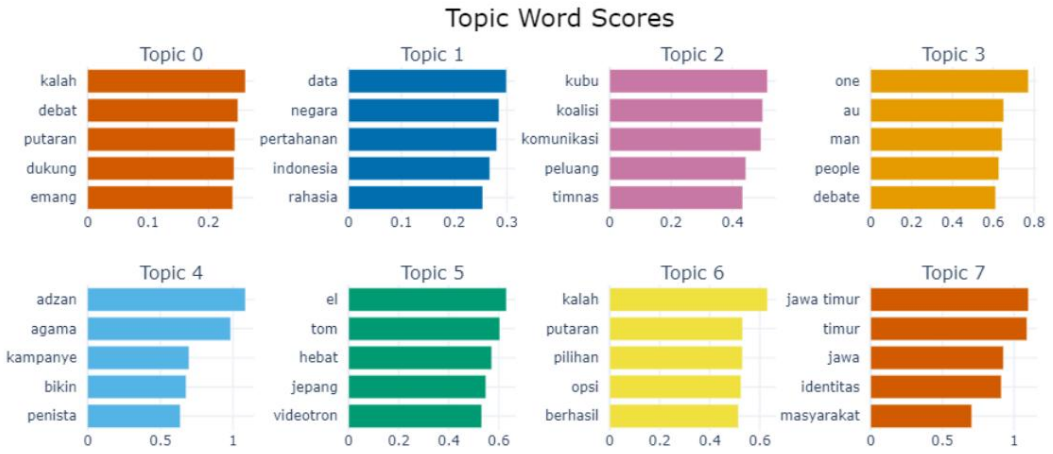


Figure 11: List of Discussion Topics Regarding Category 5 Data

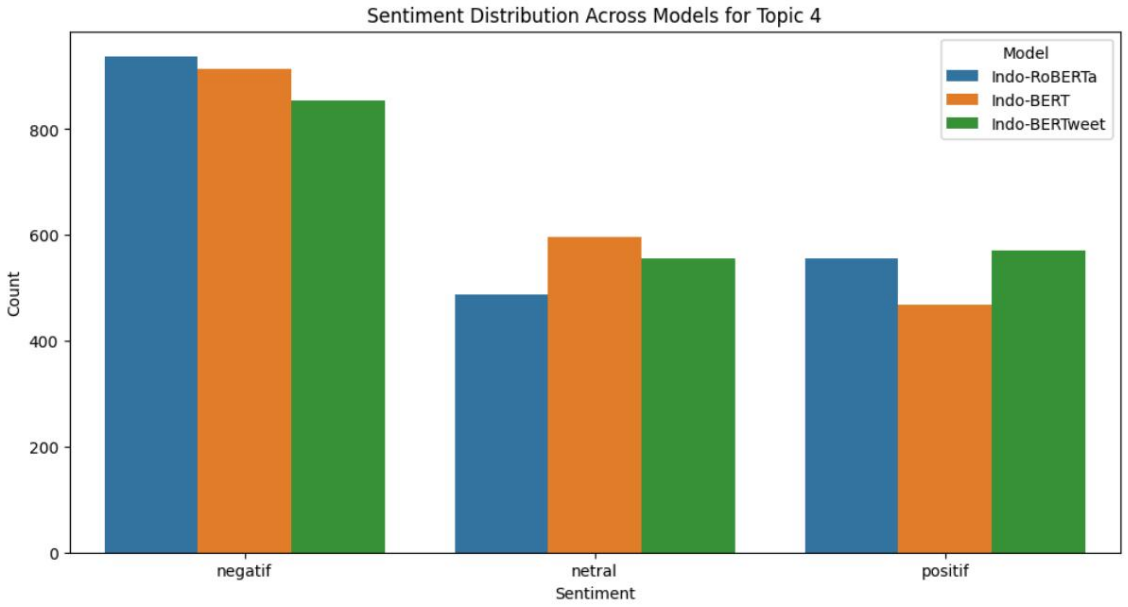


Figure 12: Distribution of Sentiment Topics Regarding Category 5 Data Regarding Topic 4

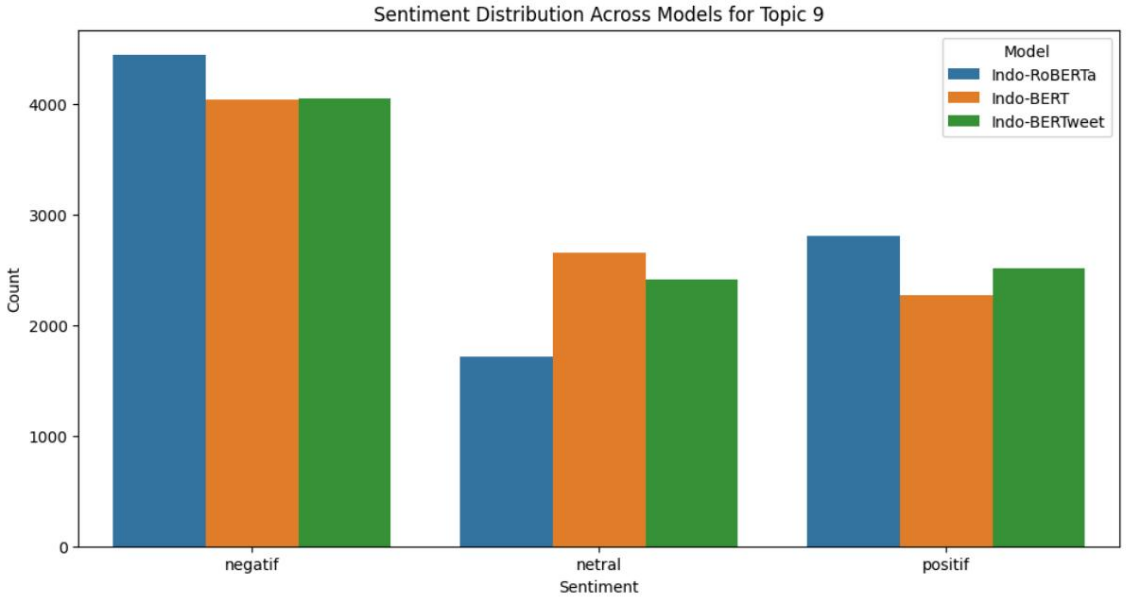


Figure 13: Distribution of Sentiment Topics Regarding Category 1 Data About Topic 9

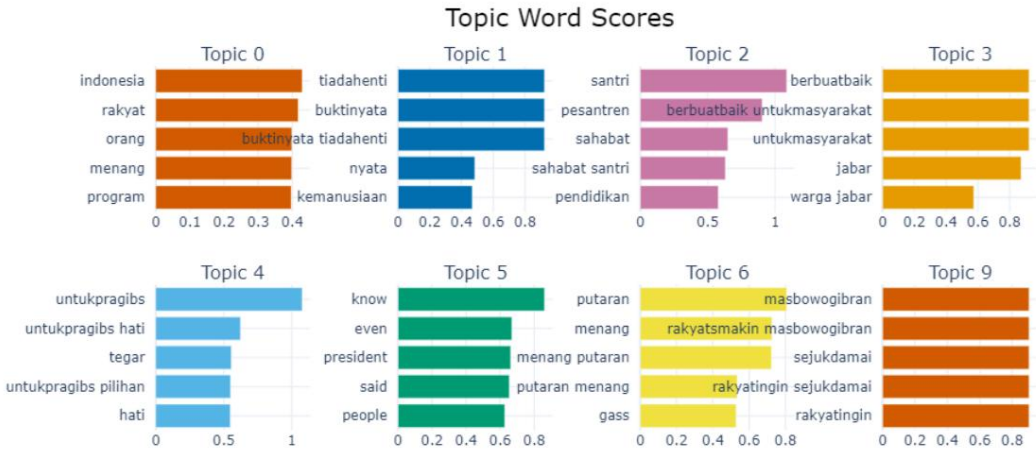


Figure 14: List of Discussion Topics Regarding Category 2 Data

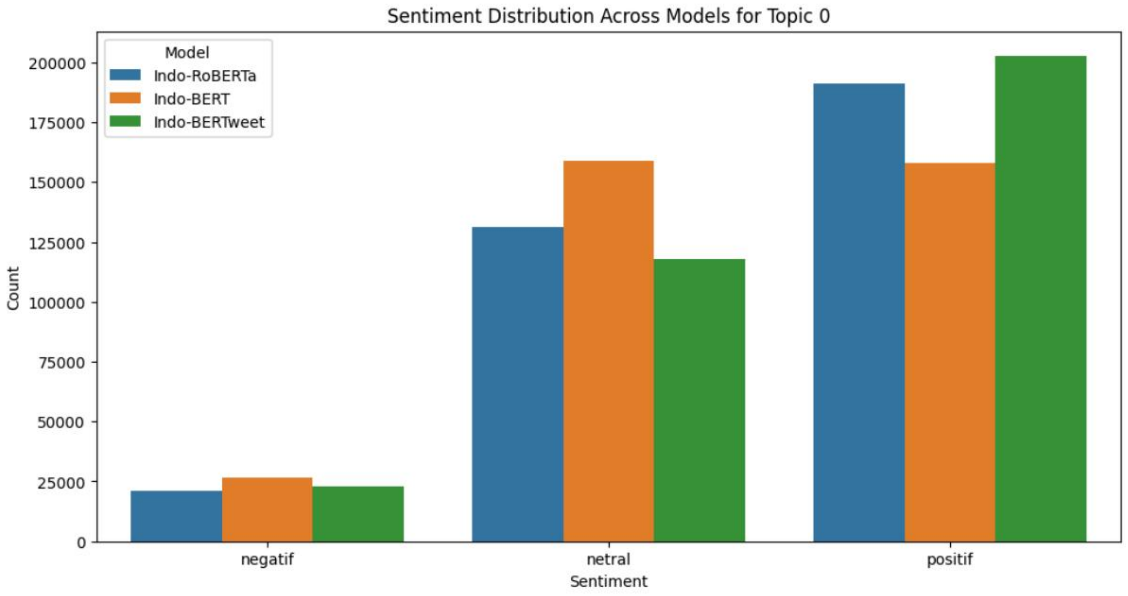


Figure 15: Distribution of Sentiment Topics Regarding Category 3 Data About Topic 0

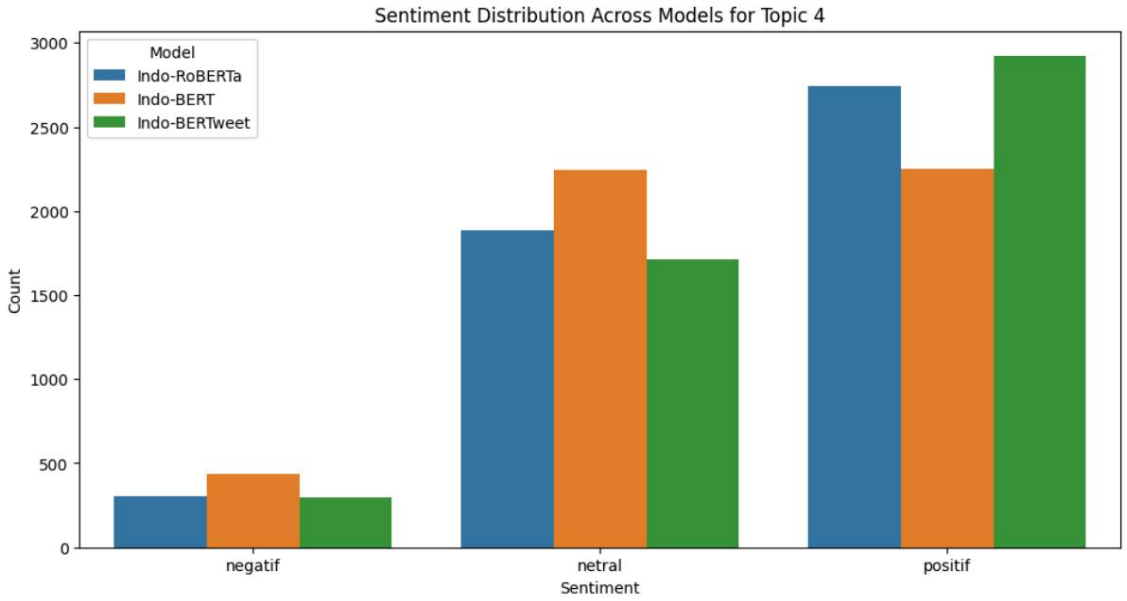


Figure 16: Distribution of Sentiment Topics Regarding Category 3 Data Regarding Topic 4

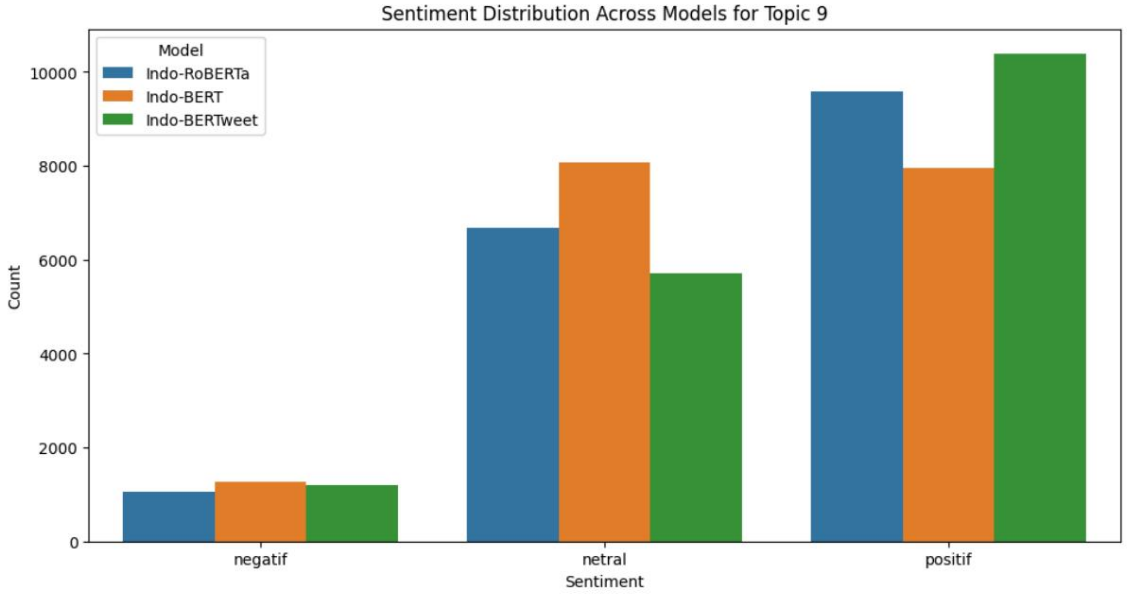


Figure 17: Distribution of Sentiment Topics Regarding Category 3 Data About Topic 9

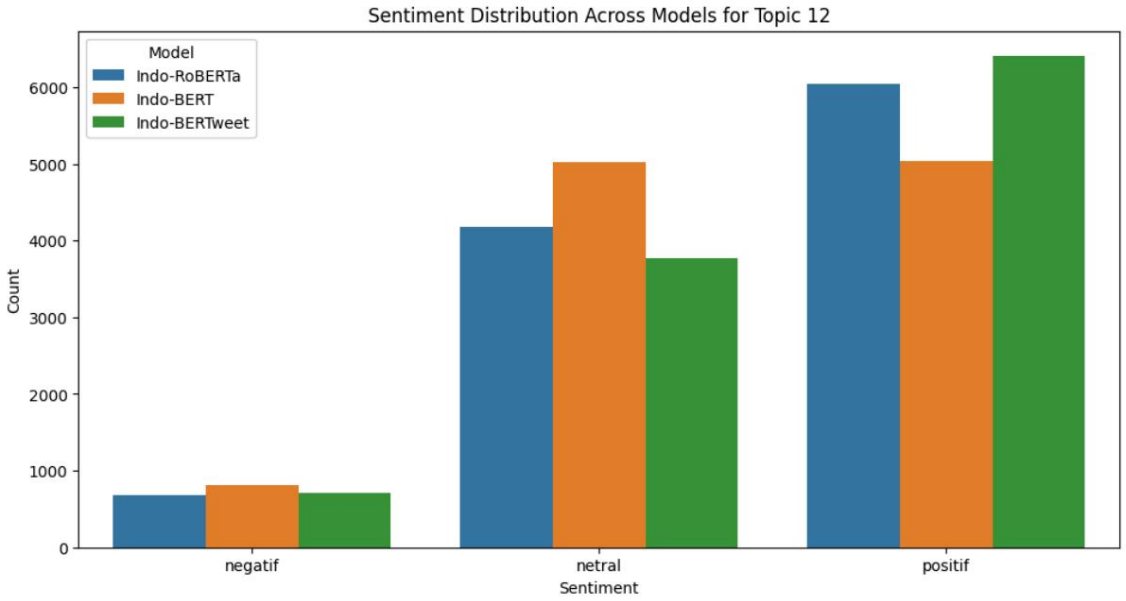


Figure 18: Distribution of Sentiment Topics Regarding Category 3 Data About Topic 12

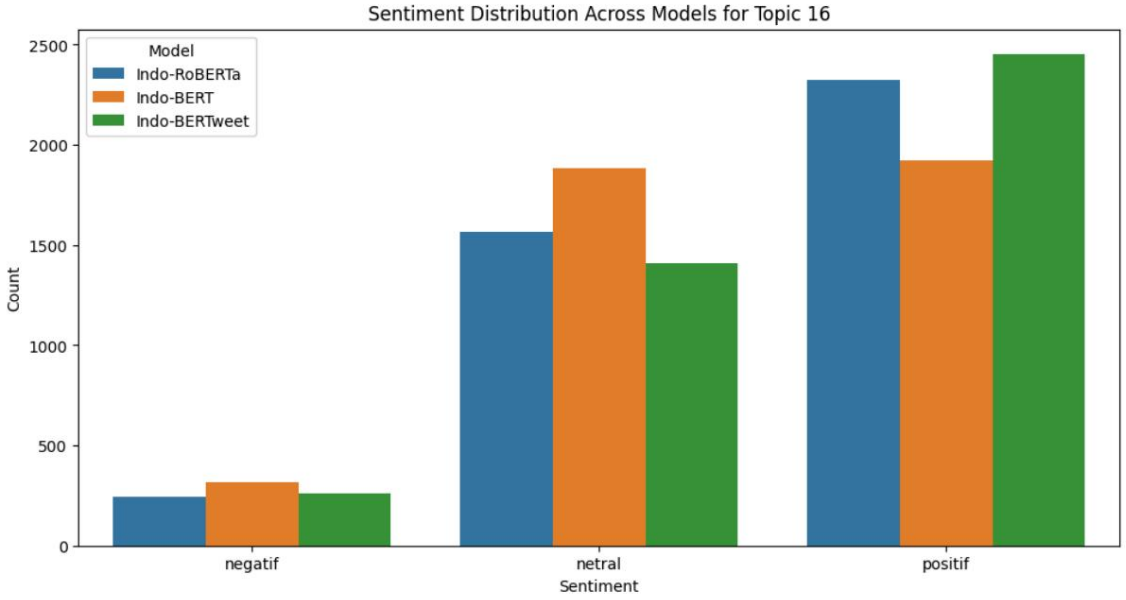


Figure 19: Distribution of Sentiment Topics Regarding Category 3 Data About Topic 16

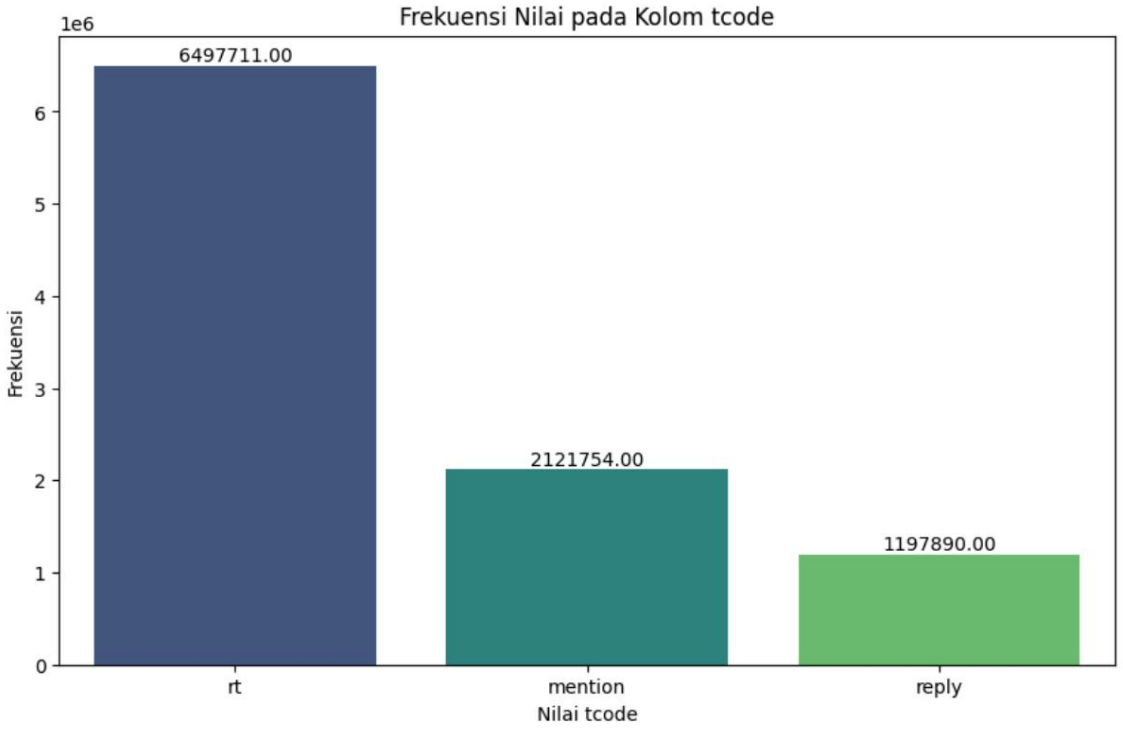
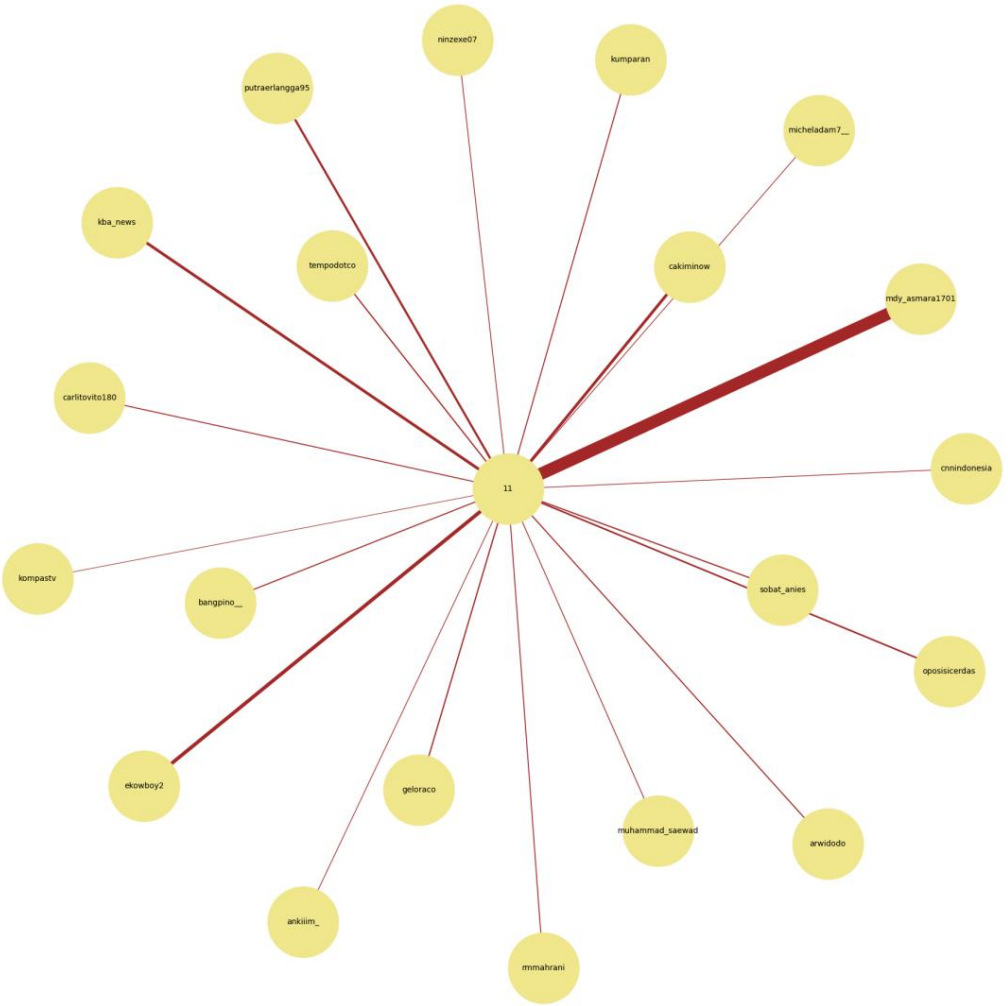
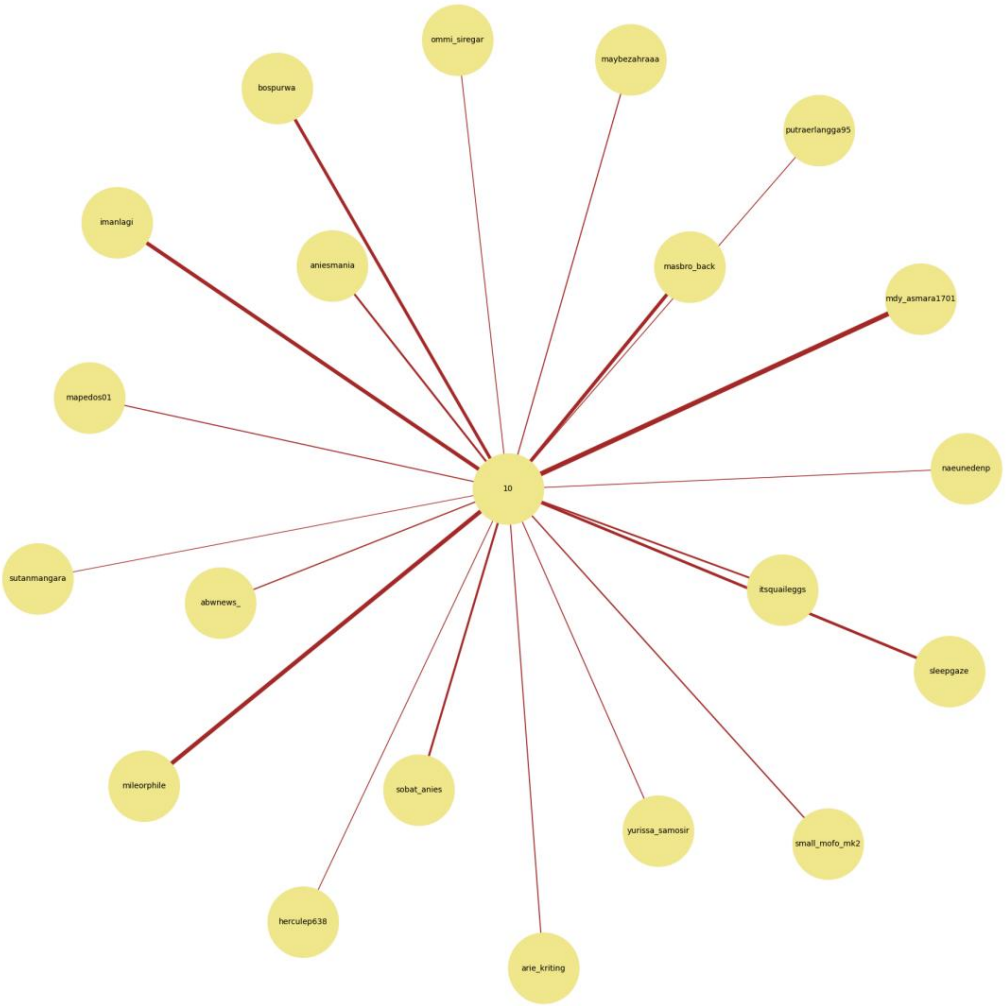
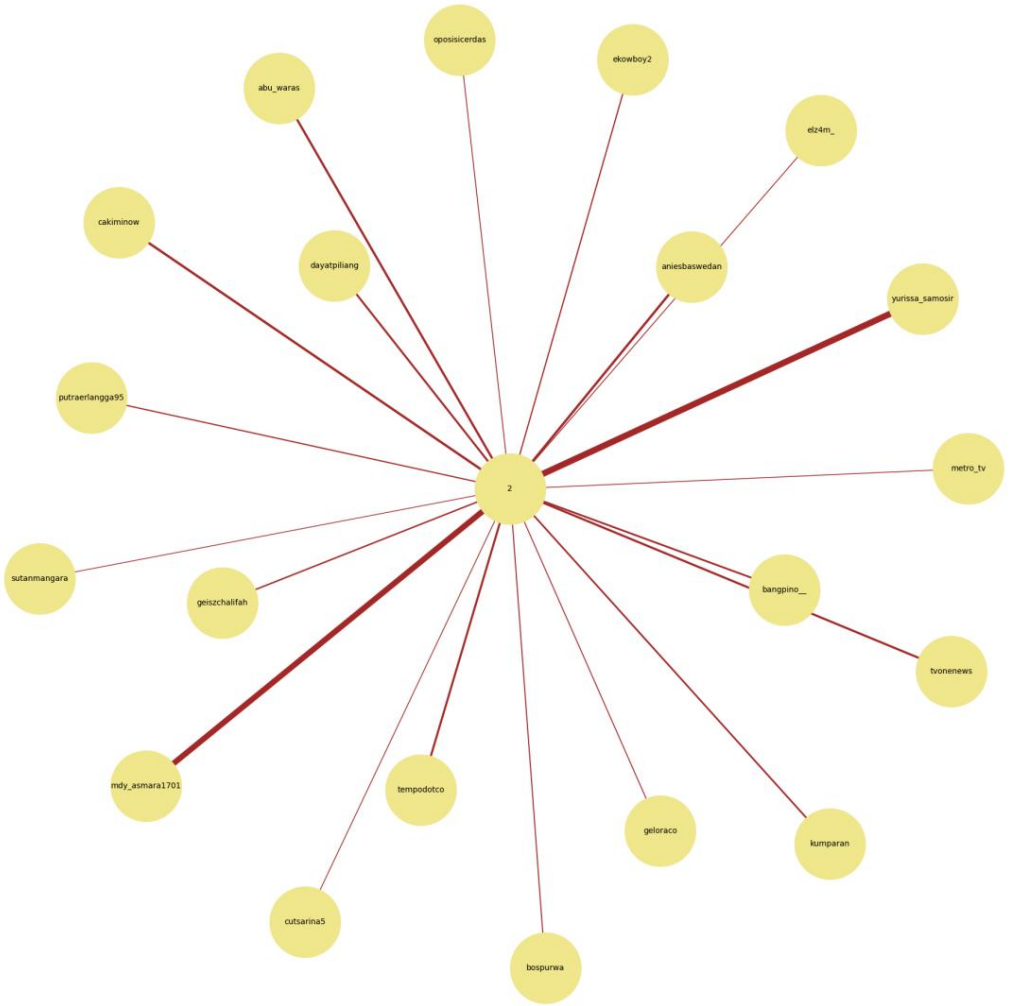


Figure 20: Frequency of Values in the tcode Column







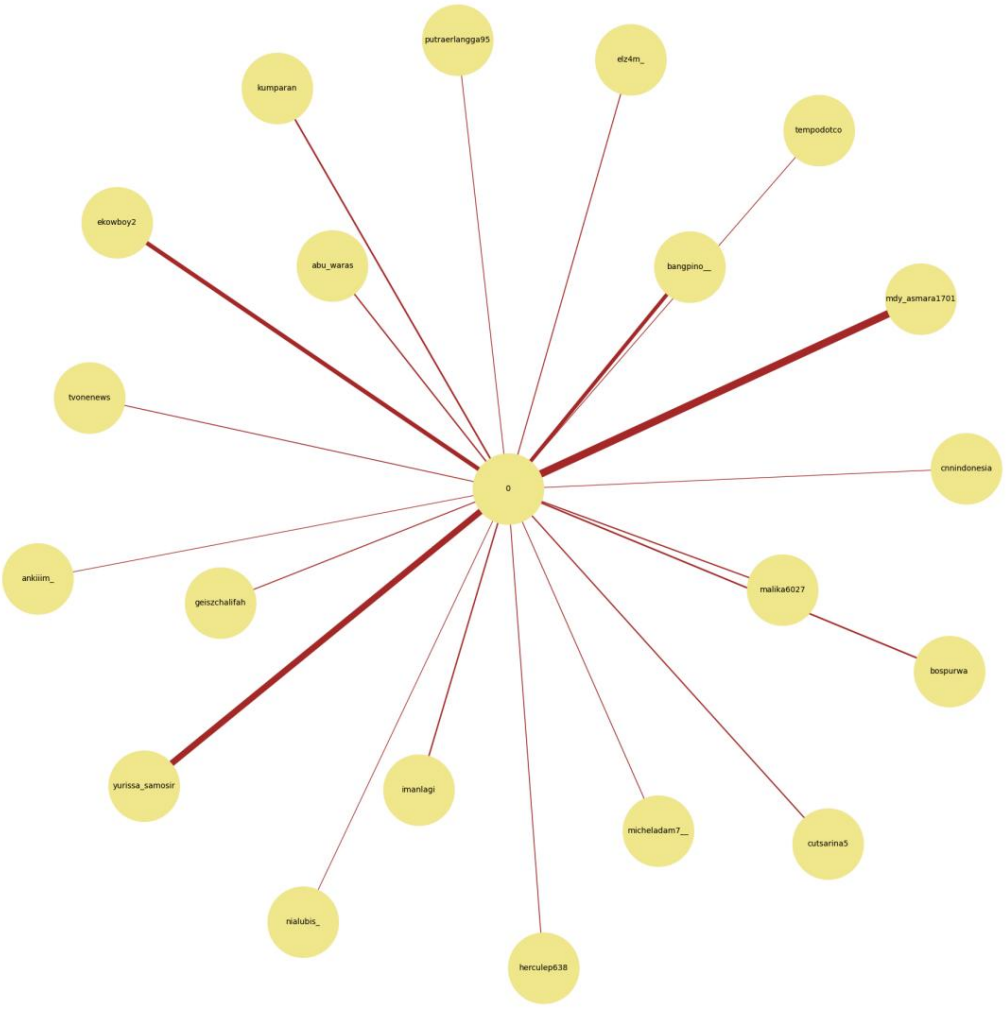
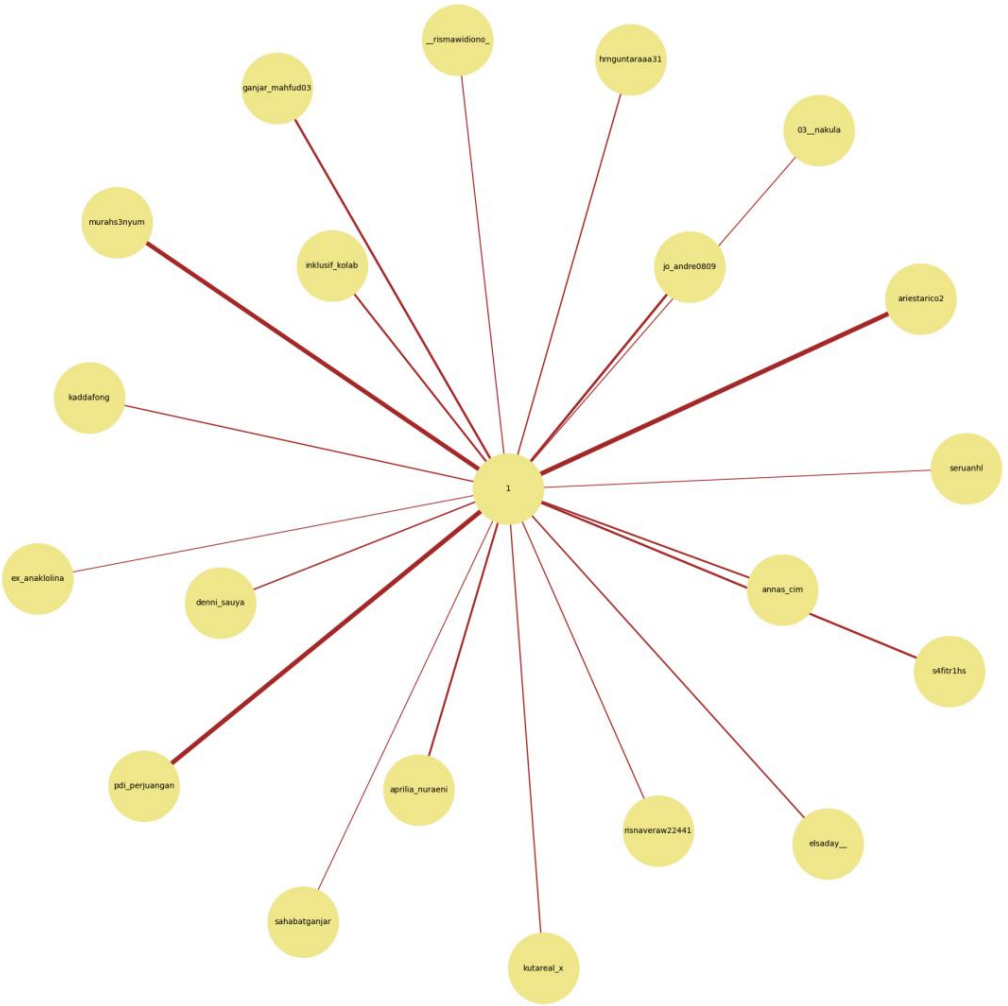
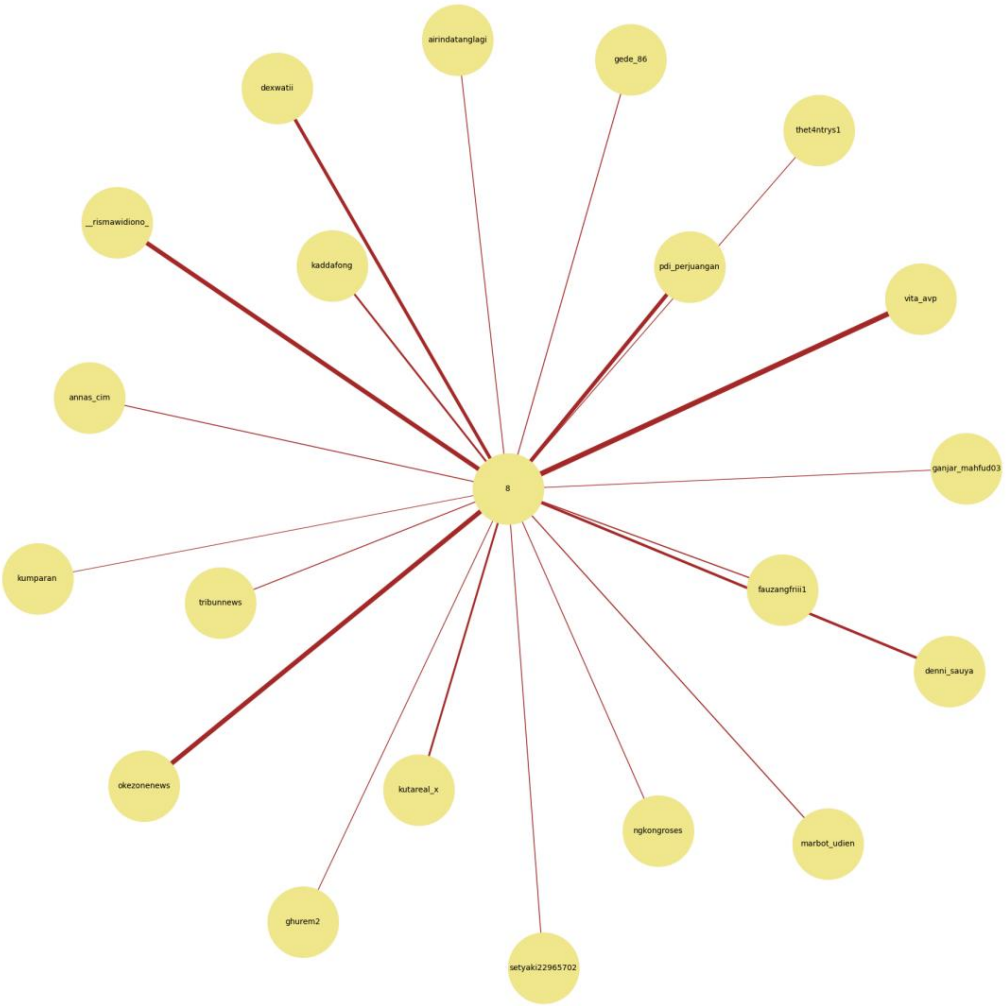
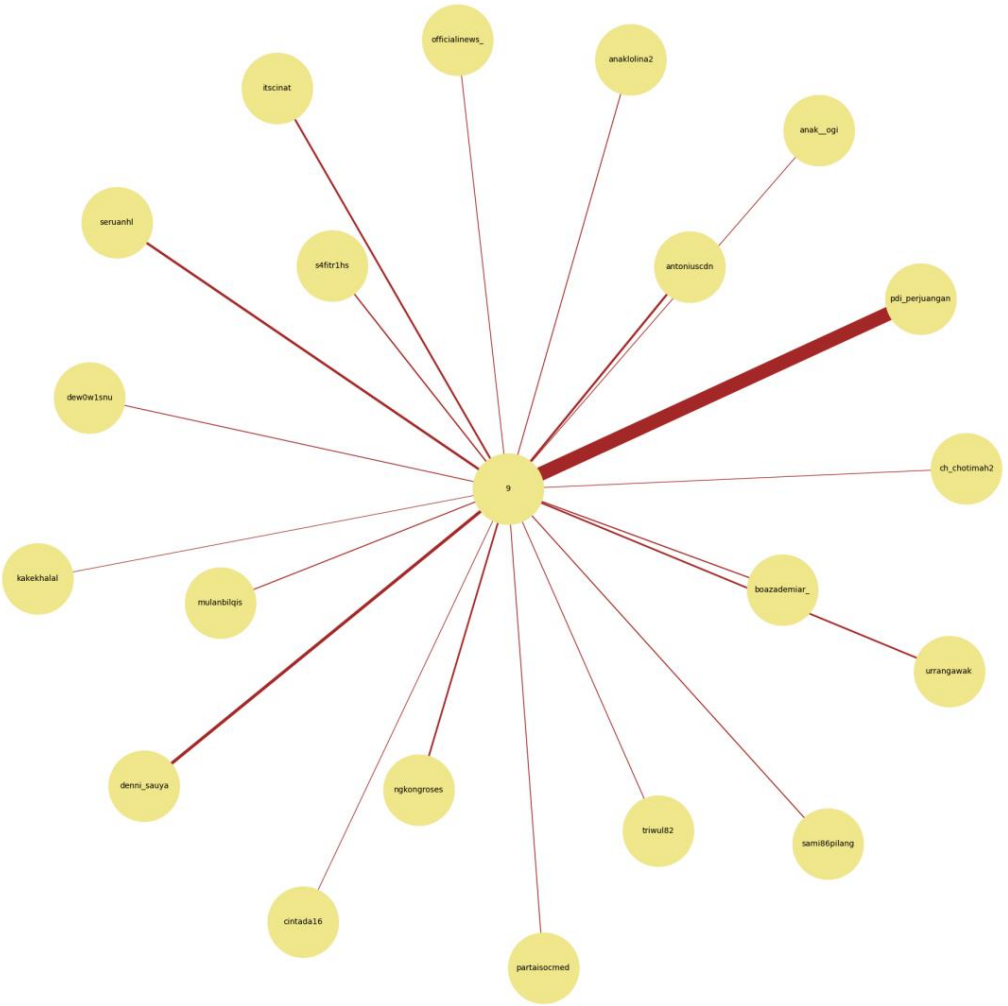


Figure 22: Category 1 Topic 11; 10; 2; 0







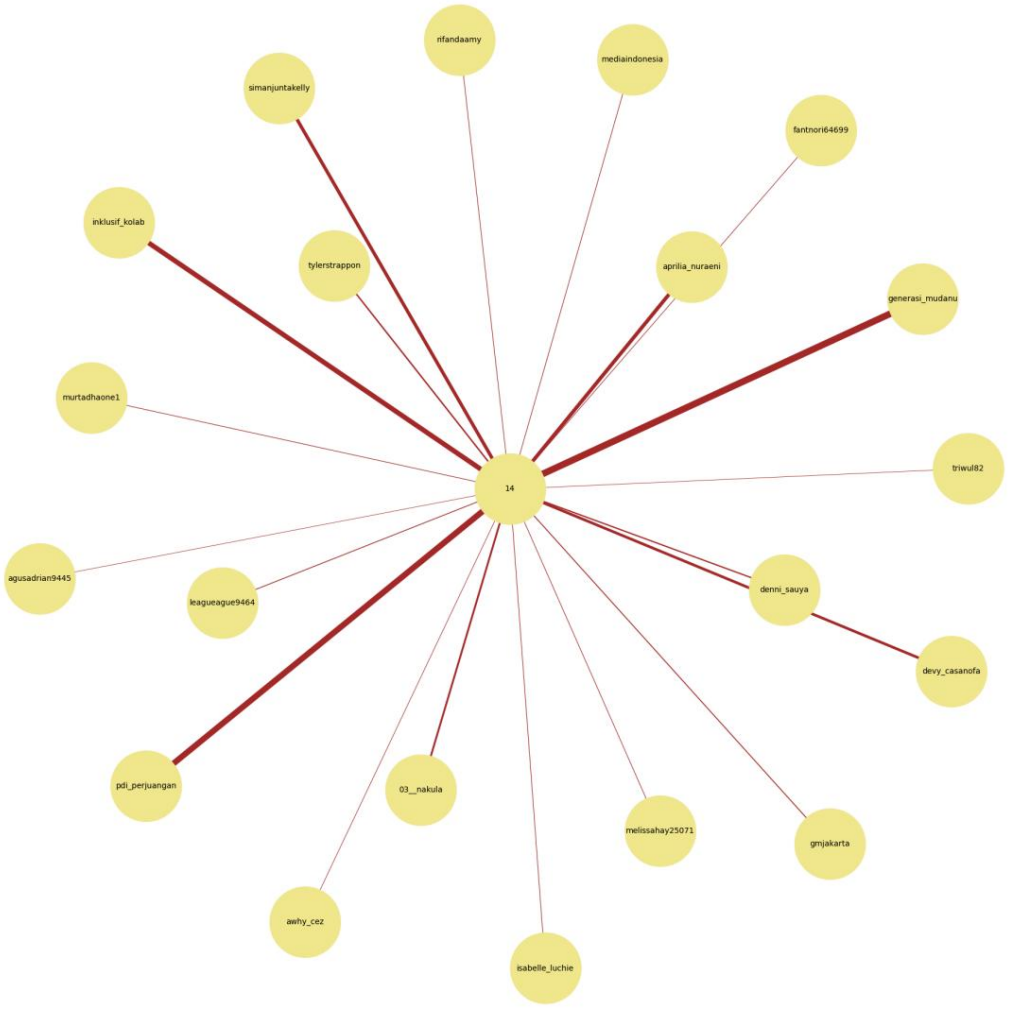


Figure 23: Category 3 Topic 1; 8; 9; 14

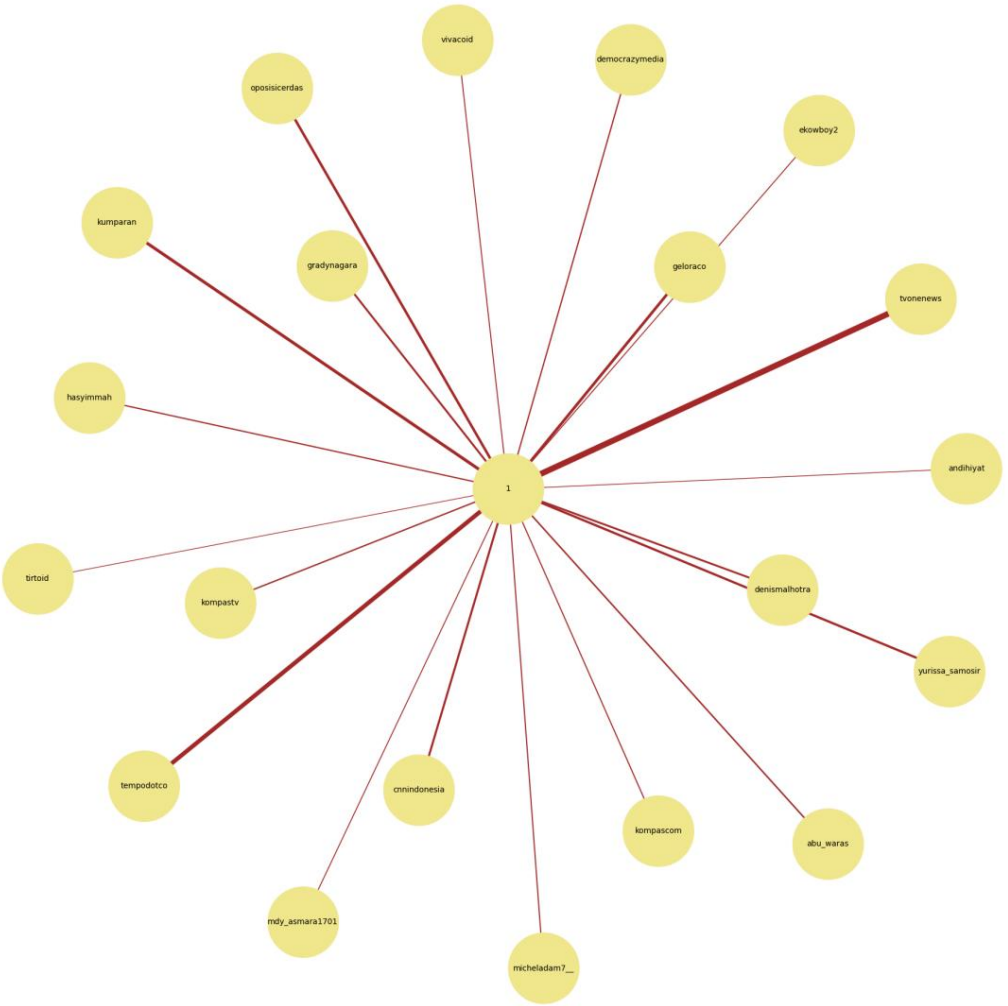


Figure 24: Category 4 Topic 1

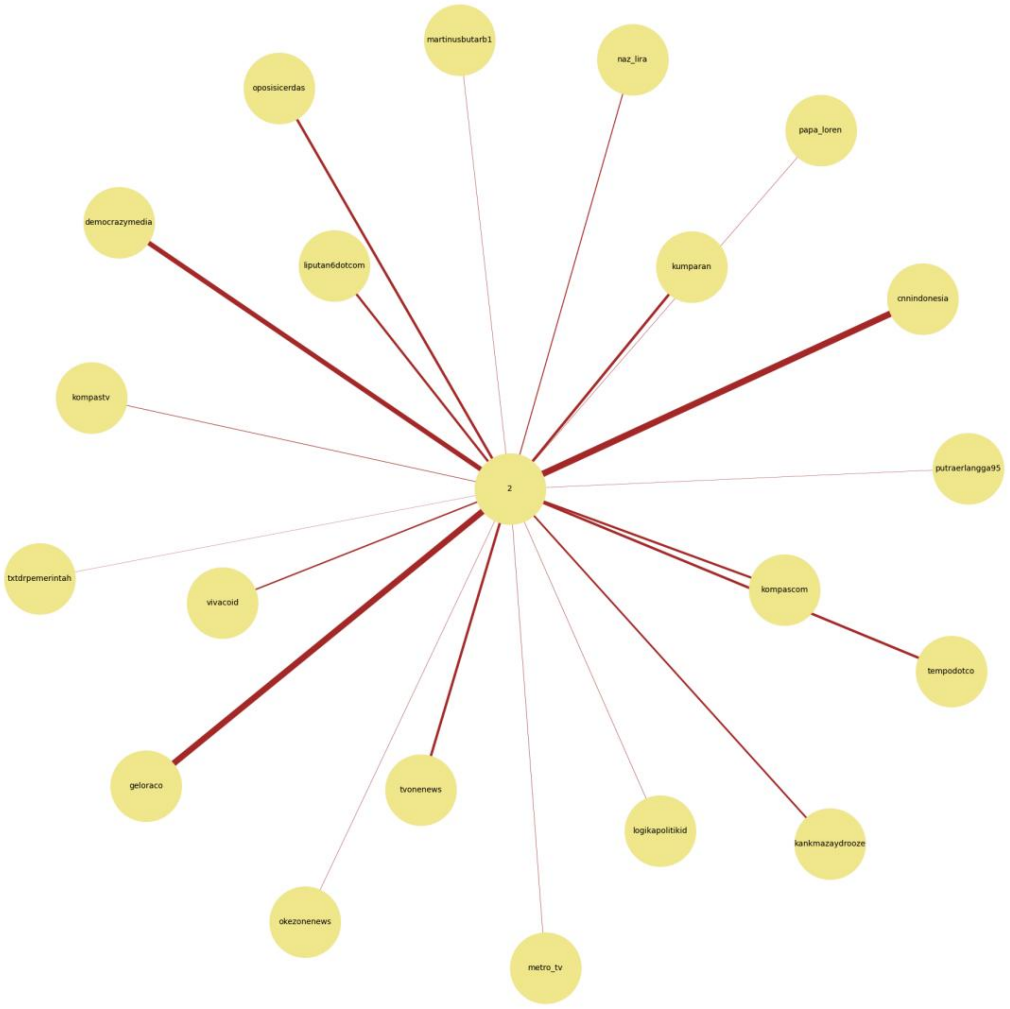


Figure 25: Category 5 Topic 2

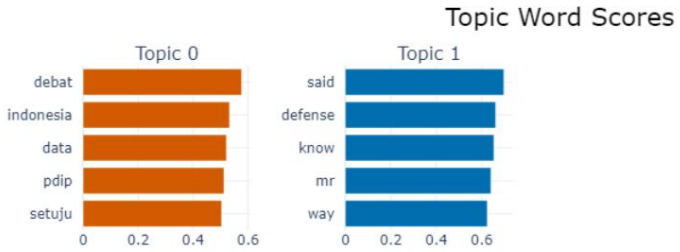


Figure 26: List of Discussion Topics Regarding Category 6 Data

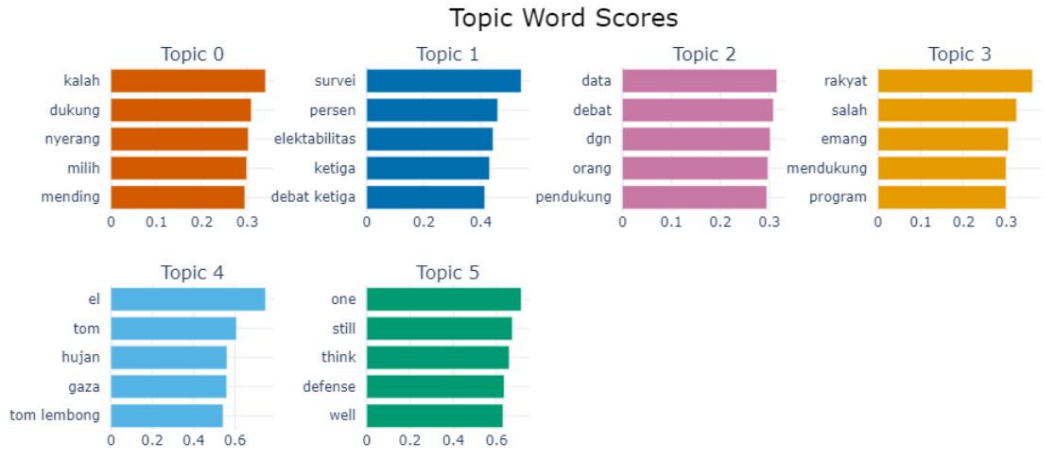


Figure 27: List of Discussion Topics Regarding Category 7 Data

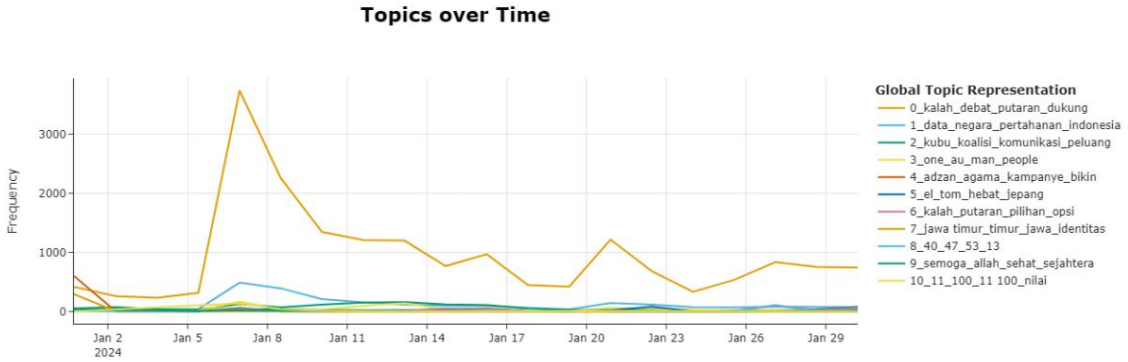


Figure 28: Graph of Discussion Topics Regarding Category 5 Data in a Period of One Month

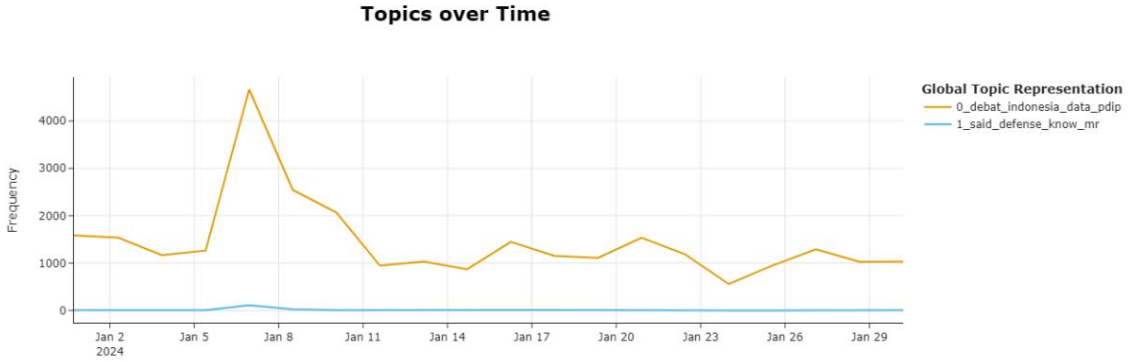


Figure 29: Graph of Discussion Topics Regarding Category 6 Data in a Period of One Month

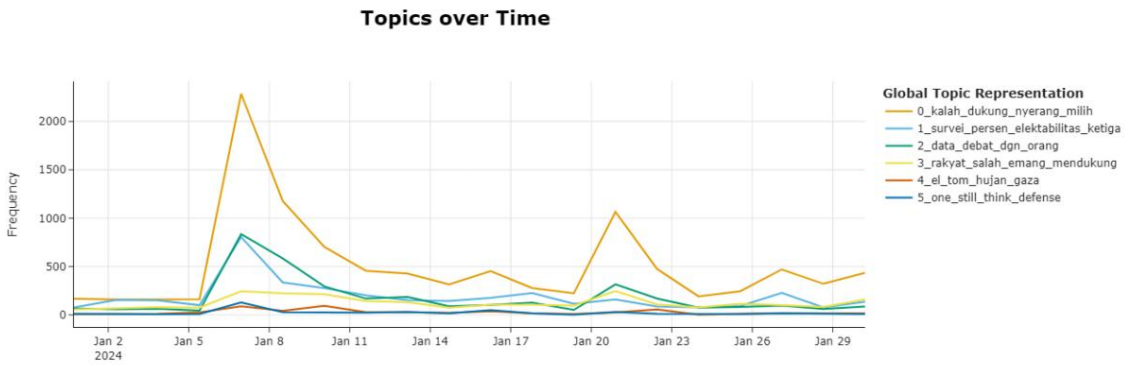


Figure 30: Graph of Discussion Topics Regarding Category 7 Data in a Period of One Month

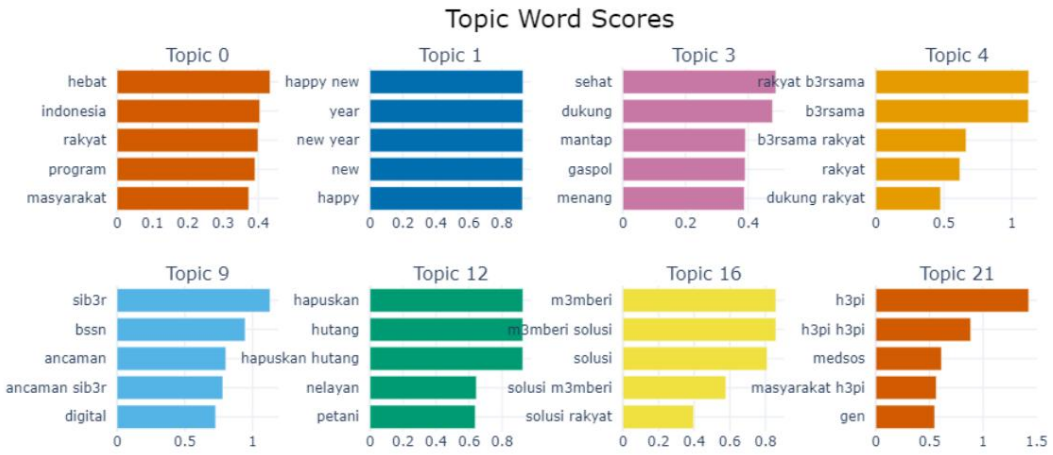


Figure 31: List of Discussion Topics Regarding Category 3 Data