The Indonesian Journal of Computer Science
www.ijcs.net
Volume 13, Issue 6, December 2024
https://doi.org/10.33022/ijcs.v13i6.4508

Analysis of General Election Campaign Topics of Candidates for President and Vice President of the Republic of Indonesia Using Lattent Dirichlet Allocation on Social Media Data

Ericko Rinanto Pratama[1], Abdullah Hasan[2], Indra Budi[3], Aris Budi Santoso[4], Prabu Kresna Putra[5]
ericko.rinanto21@ui.ac.id[1], Abdullah.Hasan@ui.ac.id[2], indra@cs.ui.ac.id[3], aris.budi@ui.ac.id[4], prabu.kresna@ui.ac.id[5]
[1,2,3,4,5]Faculty of Computer Science, Universitas Indonesia, Jakarta, Indonesia

Article Information

Received : 19 Nov 2024
Revised : 26 Nov 2024
Accepted : 30 Dec 2024

Keywords

Camapaign, Social Media, Social Media Analytics, Topic Modeling, Sentiment Analysis, Digital Campaign

Abstract

The election of the President and Vice President of the Republic of Indonesia 2024 is an important moment for the Indonesian people in determining future leaders. Social media plays a major role as a platform used to deliver campaign programs by presidential and vice presidential candidates. In conducting social media analysis, one approach that can be used is using topic modeling. Topic modeling produces output in the form of topics of conversation from a document, one of the models is Latent Dirichlet Allocation (LDA). In previous research, LDA has been widely used to search for topics of conversation on social media. This research analyzes the campaign programs of the 2024 Presidential and Vice Presidential Candidates of the Republic of Indonesia on social media using the Latent Dirichlet Allocation (LDA) method for intent classification in campaign program detection and sentiment analysis to check sentiment analysis. Data from the Twitter social media platform during the campaign period was processed and analyzed with LDA to understand the trend of campaign topics.

A. Introduction

Campaigns are strategies used in various fields to achieve a specified goal. In the context of marketing, a campaign means a series of activities that are systematically designed and organized to facilitate the exchange of certain commodities, products or services between sellers and buyers in a predetermined market with the aim of achieving regulated marketing objectives [1]. Campaigns can also be understood as organic or paid efforts to promote a brand, product or service using various advertising vehicles and other promotional tools [1].

In the political arena, on the other hand, campaigns are commonly used to support a particular political candidate, usually through the electoral process. An election campaign is the process of coordinating and organizing activities designed to influence voters' decisions in an election. Campaigns utilize a variety of political marketing tools and strategies such as mass communication tactics, advertising, lobbying, and social influence to achieve their goals [2]. Elections, recognized as important political events that determine the political direction of a country, are common in democracies around the world. Campaigns, equally important, are an integral part of these elections, influencing voters, setting political agendas and shaping discourse at national and international levels. The nature and conduct of these campaigns differ significantly based on varying cultural, historical, political, and socio-economic contexts [3].

Europe features a very diverse political landscape with different campaign styles and strategies, ranging from the highly party-centered campaigns in the UK to the French focus on individual presidential candidates [4]. Moving to Asia, countries such as India and Japan show high levels of voter turnout and party-based campaigns but differ greatly in their political climate and campaign rhetoric [5]. Oceania, with countries such as Australia and New Zealand, has strong campaign finance laws and uses both traditional and modern techniques in its campaign strategies.

In Indonesia itself, campaigns have an important meaning because they provide an opportunity for election participants to communicate their vision and mission to voters, in addition, campaigns in Indonesia have a longer time allocation compared to other stages [6]. According to Law 7 of 2017 concerning general elections, hereinafter referred to as the Election Law, the campaign stage in the 2019 elections is allocated for 3 days after the fixed list of candidates for the DPR, DPD, Provincial DPRD, Regency / City DPRD, and Candidate Pairs for President and Vice President is determined, until the start of the quiet period. In conducting a campaign, there are many resources that can be used, one of which is social media [7].

In the Merriam-Webster dictionary, social media is defined as a form of electronic communication, such as websites or social networks and microblogging, where users create online communities to share information, ideas, personal messages, and other content, such as videos. While the definition of social media in the large Indonesian dictionary explains that social media is a page or application that allows users to create and share content or engage in social networking. The rapid emergence of technological advances has made social media a common communication tool used by individuals and many people around the world.

In 2021, there are approximately 4.2 billion social media users worldwide, which represents 53.6% of the global population [8] According to Date portal, there are 160 million social media users in Indonesia, which is equivalent to 59% of the total population of Indonesia [8]. Social media utilization can be used in the campaign process both for a product campaign to a political campaign. Social media utilization also provides a channel for voters to interact directly with candidates, strengthening a more interactive and responsive democratic process. It offers an innovative way for candidates to target specific demographics and customize their messages to have maximum impact.

The purpose of the campaign process is to gain support or votes to be given by voters. However, the question arises how to analyze the campaign process on social media and what methods can be used to analyze campaigns on social media. In determining the right methodology in conducting social media analysis related to campaigns on social media, one of them can be done by using Social Media Analytics. The definition of Social Media Analytics (SMA) is a practice that involves collecting data from various social media platforms and conducting deeper analysis to generate insights for business decision making [9].

The application of SMA is very broad, ranging from detecting hate speech, analyzing social media sentiment, conducting topic analysis, and several other applications in social media analysis. [10]. However, there are several challenges faced when applying SMA, including data accuracy, data privacy, and the choice of methodology for analysis [11]. In conducting SMA, the method chosen in conducting the analysis determines the results of the analysis, therefore a methodological analysis process is needed. From these problems, a number of research questions arise as follows:

- RQ 1: What topics are there in the general election campaign program of the presidential and vice presidential candidates of the Republic of Indonesia on social media?
- RQ 2: Can the Social Media Analytics method identify topics that are discussed according to what is discussed on social media?

This research will be conducted through the following steps. First, pulling text data from Twitter social media portals and news websites. Second, pre-processing the data to produce data that is ready to use the LDA model to identify topics. Third, finding the best number of topics by using intent classification for campaign program detection and sentiment analysis to see sentiment. Fourth, apply the Latent Dirichlet Allocation (LDA) model method with topics that have been determined from the previous topic search results. Fifth, analyze the results of the LDA model to get information about the topic or campaign program for the general election of the presidential and vice presidential candidates of the Republic of Indonesia.

Another study [27] used supervised LDA, correlating topics with sentiment scores using InSet Lexicon for sentiment labels. The main difference from the previous work [27] is the use of supervised learning, which requires labeled data to predict sentiment, whereas our work uses an unsupervised approach to discover hidden topics without the need for initial sentiment labels. The previous study used InSet Lexicon for sentiment classification, while this study [27] used SVM to divide sentiment into positive, negative, and neutral categories.

This is a cutting-edge development in social media analysis for political campaigns, and future enhancements could include the integration of deep learning models such as BERT for sentiment analysis, or Transformer-based models for more accurate topic detection.

B. Research Method

We use *text mining* with the LDA algorithm to determine the dimensions of election campaign programs in text on social media, we chose the LDA algorithm because based on research on Indonesian text, the LDA algorithm has better performance than other topic modeling algorithms such as *Non-Negative Matrix Factorization* (NMF) and *Gibss Sampling Dirichlet Multinomial Mixture* (GSDMM). We then further analyze the topic modeling results of the LDA algorithm to find the dimensions of the presidential and vice presidential election campaign programs in Indonesia.

According to Blei, Ng, & Jordan [23], LDA is defined as follows:

$$p(\beta_{1:k}, \theta 1{:}\,D, z_{1:D}, w_{1:D}) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_i)\, (\prod_{n=1}^{N} p(z_{d,n}|\theta_d) p(w_{d,n}|\beta_{1:K}, z_{d,n}))$$

LDA is formally described with several notations. Topics are described by the notation β_(1:k), where β_k is the distribution over words. The topic proportion for the dth document is θ_d, where θ_(d,k) is the proportion of topic k in document d described by z_d, where z_d is the topic selection for the nth word in document d is w_d, where w_(d,n) is the nth word in document d, and is part of the fixed vocabulary. With this notation, the generative process in LDA corresponds to the following distribution of hidden and observed variables:

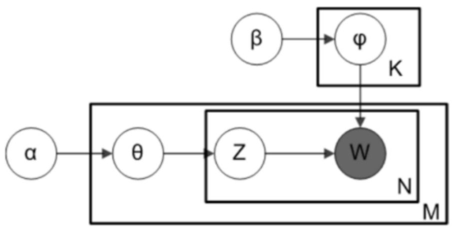

Figure 1. Plate Notation in LDA Formal Notation.
Source: [24]

The stages we undertook in carrying out this research are shown in "Figure 2". The first stage is data collection. We got the data from social media using *twitter* and news *websites*. We used the time from November 28, 2023 to February 10, 2024, taking the data based on the time of the presidential and vice presidential election campaign in Indonesia in 2024. Next, we do the pre-processing stage which we do is removing non-alphanumeric characters, equalizing the type of letter (case folding), removing punctuation marks, removing excess spaces, tokenization, removing stop words and *stemming*.

The next step is to model the topic of each review using the LDA algorithm. After getting the topic, we conducted further analysis to get the dimensions of the work programs of the presidential and vice presidential candidate pairs.

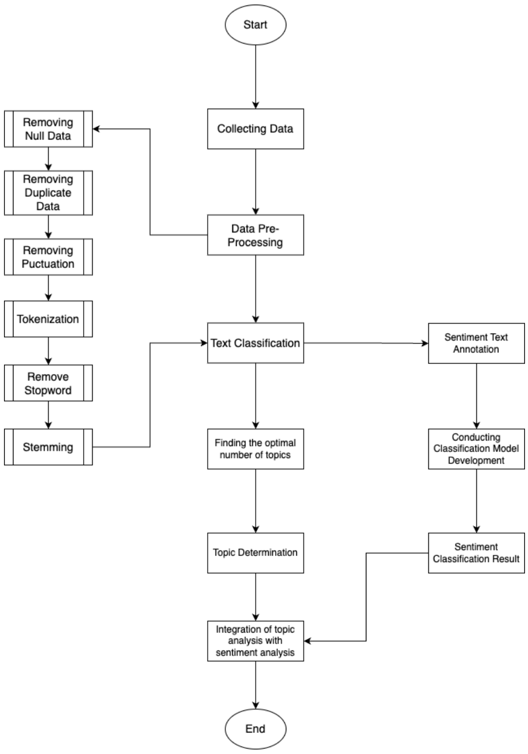

Figure 2. Research Flow

1. Collecting Data
   In the process of collecting data, data is obtained by pulling from *twitter* social media and news websites*.* From the data collected in csv format. With available time from 2023-11-28 to 2024-02-10. The csv data has *created at, id_str, full_text, quote_count, reply_count, favorite_count, lang, conversation_id_str, username,* and *tweet_url* columns*.*

2. Data Preprocessing
   Data preprocessing is done with several stages, the first of which is
   - *Removing Null Data*: The process of removing or eliminating data that has no or invalid values, such as NaN, None, or NaT, from the dataset. This process can help improve the quality and accuracy of data analysis.
   - *Removing Duplicates*: this process involves *removing* documents that have similarities with other documents in the corpus. This process is useful to improve the performance of *text mining* in the next process.
   - *Removing Punctuation*: the process of *removing punctuation* from the text to be analyzed. This process aims to reduce *noise* and make it easier to identify words that are relevant to a particular topic.
   - *Case folding*: this process converts all letters starting from "a" to "z" in the document into lowercase letters. Not all documents are consistently capitalized. Therefore, *case folding* is used to convert all text in the document into lowercase letters.

3. Transformed Data
   At this stage the results obtained from data pre-processing will be subjected to a data transformation process. Data transformation includes reducing

the number of words that are not meaningful in the *stop words* process and converting words into their basic words with the *stemming* process. The explanation of the text transformation stage is as follows:

- *Removing stop words*: *stop words* are common words that usually appear in large numbers and are considered meaningless. *Stop words* have little influence in a sentence/text, so they need to be filtered to improve the performance of the text *mining* process. Common words (such as "and", "or", "at") that do not add important information in text analysis are removed.
- *Stemming*: is a process used to improve the performance of *information retrieval* by transforming words in text documents into their basic word form. The *stemming* process in Indonesian text is more complicated/complex because there are a variety of affixes that must be removed to get the *root word of* a word. For example: "membetulkan" becomes "betul", "berpegangan" becomes "pegangan".
- Performing vectorization, at the vectorization stage changes the text document into a vector. Changing the text into a vector aims to adjust to the input that the *machine learning* model will use in identifying patterns.

4. Performing Text Classification
   At the text classification stage, from the text that has been tokenized, the number of occurrences of words is seen, then a sentence or text is selected based on words related to the election program. Such as words or texts that have the word "program" or texts that provide information about words related to education, health, economy, infrastructure, security and other programs. After a text data has been classified which ones talk about the program and which ones do not, then the next analysis is carried out, namely building sentiment analysis and campaign topics using LDA.

5. Sentiment Analysis
   In the sentiment analysis stage, several stages are carried out such as annotating text data which divides into three categories namely "neutral", "positive", and "negative". After annotating the data that is really *clean,* the next step is to build a sentiment analysis model to classify the sentiment model. The model used in the classification is *support vector machine* (SVM). In building the SVM model, the data that has been annotated is used to perform learning, and the next stage is to predict the *testing* data that will be used.

6. Find the Best Corpus
   To get the appropriate number of topics, the *coherence* value is used as a reference for determining the optimal number of topics to be used in the next process. *Coherence* measures the quality of the topics generated by the topic model by looking at how well the words in a topic match each other. *The coherence* value in the LDA model is a measure that shows how

consistent and coherent the topics generated by the model are. A high *coherence* value indicates that the topics are easily understood by humans and have a clear meaning.

7. Determination of *Topic Modeling* Dimensions
   After obtaining the best number of dimensions, the next step is to build an LDA model in the process of identifying the dimensions of the work program of presidential and vice presidential candidates during the election. The following are the stages of this process:
   - Search for representative words for each topic generated by the LDA model. This step can be done using python program code utilizing the genism library, which provides a function to retrieve the top words from each topic.
   - Association of representative words with specific topics or dimensions that fit the context of the data. This step was done manually based on the author's intuition, taking into account the meaning and semantic relationship between the words.
   - Determining the coherent value uses statistics in the text corpus to determine how often two words appear in the same context, and then normalizes the coherent value.

C. Result and Discussion

1. Sentiment Analysis

   In this study, a text-based dataset is the focal point of our analysis, which contains the full text of tweets posted by individuals related to the election campaign. The purpose of this analysis is to explore the nuances of sentiment contained in the tweets to understand more about public perceptions and reactions to the ongoing election campaign.

   In text processing, a number of steps have been applied including removal of emojis, symbols, and punctuation marks, conversion of all text to lowercase, converting slang terms to common forms, removing stop words, performing stemming, performing POS tagging, and performing vectorization.

   The 'kernel' parameter in SVM specifies the kernel function used to transform the input into a high-dimensional space so that the data can be linearly separated. The 'linear' option uses a linear kernel, while 'rbf' (Radial Basis Function) is a non-linear kernel suitable for non-linear data. Meanwhile, the 'C' parameter is a regularization parameter that strikes a balance between obtaining a larger margin and avoiding misclassification. Lower values of 'C' result in wider margins with higher error tolerance, while higher values create narrower margins with lower classification errors. The hyperparameter tunning results show that the best model is using C=10 and the rbf kernel, which achieves 70% accuracy.

```
--- Tuning SVM ---
Best Parameters for SVM: {'C': 10, 'kernel': 'rbf'}
              precision    recall  f1-score   support

    negative       0.63      0.74      0.68        89
     neutral       0.76      0.86      0.81       109
    positive       0.70      0.18      0.29        39

    accuracy                           0.70       237
   macro avg       0.70      0.59      0.59       237
weighted avg       0.70      0.70      0.68       237
```

Figure 3. SVM Model Tuning Result

The 'max_depth' parameter in the decision tree sets the maximum depth that the tree can reach. A value of 'None' allows the tree to grow without any depth limit, which can increase the risk of overfitting. Setting a maximum depth such as 10 or 20 can help reduce this risk. Meanwhile, 'min_samples_split' specifies the minimum number of samples needed to split internal nodes. If the node has fewer samples than this value, the split will not be performed, with higher settings helping to prevent overfitting. Based on the hyperparameter tunning process, the best model was obtained with a 'min_samples_split' of 2 and recorded an accuracy of 68%.

```
--- Tuning Decision Tree ---
Best Parameters for Decision Tree: {'max_depth': None, 'min_samples_split': 2}
              precision    recall  f1-score   support

    negative       0.64      0.67      0.66        89
     neutral       0.74      0.82      0.77       109
    positive       0.50      0.28      0.36        39

    accuracy                           0.68       237
   macro avg       0.62      0.59      0.60       237
weighted avg       0.66      0.68      0.66       237
```

Figure 4. Decision Tree Model Tuning Results

The 'n_estimators' parameter indicates the number of trees present in a forest in a random forest model, where more trees can usually improve model performance but also extend computation time. Meanwhile, 'max_features' defines the number of features considered when dividing each node, with the options 'auto' using all available features, 'sqrt' using the square root of the total number of features, and 'log2' using the base two logarithm of the number of features, which impacts the diversity between trees in the model. Based on the hyperparameter tunning process, the best combination was found by using 'max_features' of type 'log2' and 'n_estimators' of 100, resulting in an accuracy of 71%.

```
Best Parameters for Random Forest: {'max_features': 'log2', 'n_estimators': 100}
              precision    recall  f1-score   support

    negative       0.66      0.75      0.70        89
     neutral       0.74      0.88      0.81       109
    positive       0.83      0.13      0.22        39

    accuracy                           0.71       237
   macro avg       0.74      0.59      0.58       237
weighted avg       0.73      0.71      0.67       237
```

Figure 5. Random Forest Model Tuning Results

The 'alpha' parameter is a smoothing parameter in probability estimation that aims to avoid the problem of division by zero. Higher values of 'alpha' provide a stronger smoothing effect, which is especially useful when the data has features that vary in frequency. Through the hyperparameter tunning process, it was found that the model with an 'alpha' setting of 1.0 produced the best performance, achieving 74% accuracy.

```
--- Tuning Naive Bayes ---
Best Parameters for Naive Bayes: {'alpha': 1.0}
              precision    recall  f1-score   support

    negative       0.68      0.80      0.73        89
     neutral       0.83      0.83      0.83       109
    positive       0.58      0.36      0.44        39

    accuracy                           0.74       237
   macro avg       0.70      0.66      0.67       237
weighted avg       0.73      0.74      0.73       237
```

Figure 6. Naïve Bayes Model Tuning Results

The 'hidden_layer_sizes' parameter in the neural network determines the size and number of hidden layers, where a configuration such as (30,) indicates one hidden layer with 30 neurons, while (30,20) signifies two hidden layers with 30 neurons in the first layer and 20 in the second. The 'alpha' parameter, as part of the L2 regularization, helps reduce overfitting by penalizing weights that are too large. After searching for the optimal hyperparameters, it was found that using 'alpha' of 0.001 and 'hidden_layer_sizes' of 50 neurons resulted in an accuracy of 71%.

```
--- Tuning Neural Network ---
Best Parameters for Neural Network: {'alpha': 0.001, 'hidden_layer_sizes': (50,)}
              precision    recall  f1-score   support

    negative       0.67      0.72      0.70        89
     neutral       0.75      0.82      0.78       109
    positive       0.65      0.38      0.48        39

    accuracy                           0.71       237
   macro avg       0.69      0.64      0.65       237
weighted avg       0.70      0.71      0.70       237
```

Figure 7. Neural Network Model Tuning Results

The 'n_neighbors' parameter in the KNN model specifies the number of nearest neighbors used for classification, where using more neighbors will usually smooth the decision boundary, reduce noise but may reduce sensitivity to the actual data. The 'weights' parameter determines how the neighbors are weighted, with the 'uniform' option giving equal weight to all neighbors, while 'distance' gives greater weight to closer neighbors. Through hyperparameter tunning, the optimal KNN model was found using 'n_neighbors' of 3 and distance-based 'weights', resulting in 63% accuracy.

```
--- Tuning KNN ---
Best Parameters for KNN: {'n_neighbors': 3, 'weights': 'distance'}
              precision    recall  f1-score   support

    negative       0.62      0.54      0.58        89
     neutral       0.62      0.87      0.73       109
    positive       0.75      0.15      0.26        39

    accuracy                           0.63       237
   macro avg       0.67      0.52      0.52       237
weighted avg       0.64      0.63      0.59       237
```

Figure 8. KNN Model Tuning Results

2. Determination of *Topic Modeling* Dimensions
   The graph above shows the coherence score obtained from the topic model based on the number of different topics. Coherence score is a metric used to assess how meaningful or consistent the topics generated by the model are. In this graph, it can be seen that when the number of topics is 3, the model achieves the highest coherence score of around 0.635. This shows that by using three topics, the model can generate the most coherent and meaningful group of topics compared to the other number of topics. In contrast, when the number of topics is 4, the coherence score drops dramatically, indicating that the model with four topics is less effective in producing coherent topics. The coherence score then increased again at five topics, but not as high as at three topics. Thus, in this context, three topics is the optimal number to achieve the highest coherence score in topic modeling.

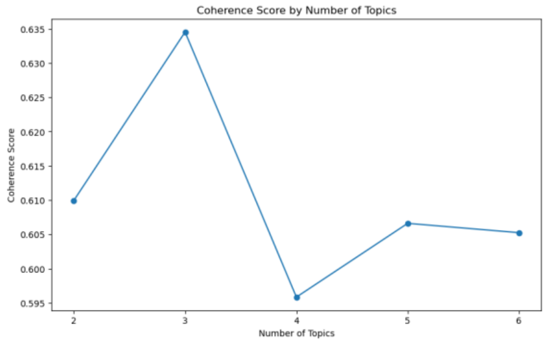

Figure 9. Coherent Value of Number of Topics

```
Best number of topics: 3.0
Topic 0:
program kampanye pemilu gibran gratis prabowo makan prabowogibran tim amin
Topic 1:
program ganjar kampanye pemilu anies 2024 bansos nomor presiden urut
Topic 2:
anies kampanye yg program aja desak gak ga ya imin
```

Figure 10. Top Three Topics

Figure 9 shows the result of the topic model with the best number of topics, which is three topics. Each topic contains key words that reflect the main theme of the topic.

Topic 0: Key words such as "campaign program," "election," "gibran," "gratis," "prabowo," "makan," and "team" indicate a focus on the election campaign program involving Gibran and Prabowo, with this topic possibly addressing campaign activities, political appointments, and election-related events.

Topic 1: Key words such as "program," "ganjar," "campaign," "election," "anies," "2024," "social assistance," "number," and "president" indicate a discussion of Ganjar and Anies' campaign programs for the 2024 election, with an emphasis on social assistance (bansos) and the presidential candidates' numbers.

Topic 2: Key words such as "anies," "campaign," "program," "aja," "urge," "pak," "ga," and "irin" highlight topics related to Anies' campaign, focusing on urgent programs and reactions from the public or certain parties.

Figure 11 shows a histogram of the most frequently occurring words in Topic 0 based on the topic modeling results. These words reflect the main themes of the topic. The word "program" has the highest frequency, indicating that it is the most frequently occurring word and may be the main focus in this topic. The words "campaign" and "election" also have high frequencies, indicating that discussions often revolve around election campaign activities. Words such as "gibran," "gratis," "prabowo," "makan," "prabowogibran," "tim," and "amen" also appeared with significant frequency, suggesting that this topic covers various aspects related to political campaigns involving figures such as Gibran and Prabowo, and their various campaign programs or promises. This analysis helps identify the main focus and important elements in the topic being discussed.

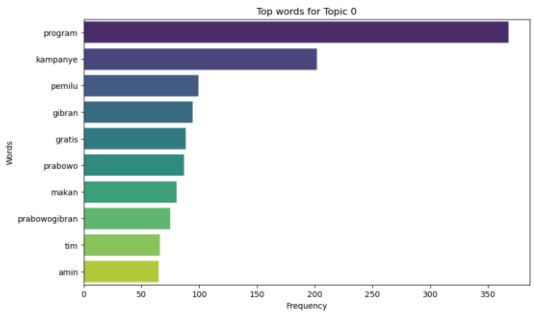

Figure 11. Topic 0

Figure 12 shows a histogram of the most frequently occurring words in Topic 1 based on the topic modeling results. The word "program" has the highest frequency, indicating that it is the most dominant word in this topic. The words "ganjar" and "campaign" also have high frequencies, indicating that this topic is mostly about campaign programs related to Ganjar. The words "election," "anies," "2024," "social assistance," "number," "president," and "sequence" appeared with significant frequency, indicating that discussions in this topic also involved the 2024 election, social assistance programs, as well as the sequence numbers of presidential candidates, particularly in relation to Ganjar and Anies. This analysis helped to identify the key elements and main focus in the topics discussed, namely political campaigns and election-related programs.

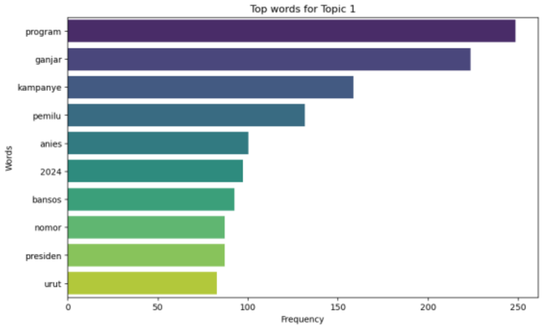

Figure 12. Topic 1

Figure 13 shows a histogram of the most frequent words in Topic 2 based on the topic modeling results. The word "anies" has the highest frequency, indicating that it is the most dominant word in this topic. The word "campaign" also has a very high frequency, indicating that much of the discussion revolves around Anies' campaign. The words "yg," "program," "aja," "desak," "gak," "ga," "ya," and "irin" also appeared with significant frequency. This suggests that this topic is mostly about the campaign programs run by Anies, with some words indicating a discussion or debate about the urgency and effectiveness of these programs. This analysis helps identify the main focus of the topic, which is Anies' campaign and the various programs and related issues that emerged during the campaign.

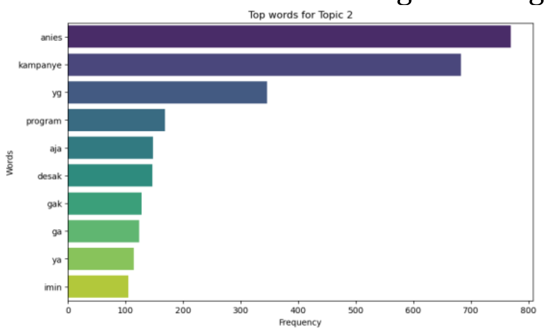

Figure 13. Topic 2

D. Conclusion

In this research, sentiment analysis and topic modeling have been conducted to understand the dynamics of election campaigns. The Naïve Bayes model was selected as the best model for sentiment analysis because it showed the highest accuracy rate among the other five models with a value of 74%. Furthermore, in the aspect of topic modeling, the model used managed to achieve a satisfactory coherence value of 3. One of the topics successfully identified in the context of the election campaign was related to the provision of free meals. This research highlights the importance of analytical techniques in understanding and interpreting data related to election campaigns, particularly through the use of the Naïve Bayes method for sentiment analysis and topic modeling techniques to uncover specific aspects of focus in the campaign.

E. References

[1] K. Clow and D. Baack, Integrated Advertising, Promotion, and Marketing Communications: Global Edition. 2022.

[2] M. P. Wattenberg, Parties Without Partisans: Political Change in Advanced Industrial Democracies: Political Change in Advanced Industrial Democracies. OUP Oxford. 2002.

[3] S. E. Scarrow, "Beyond Party Members: Changing Approaches to Partisan Mobilization," Comp Polit, 2015.

[4] D. M. Farrell and P. Webb, "Political Parties as Campaign Organizations," in Parties Without Partisans, 2003. doi: 10.1093/0199253099.003.0006.

[5] F. C. Harris, " Strength in Numbers? The Political Mobilization of Racial and Ethnic Minorities . Jan E. Leighley ," J Polit, vol. 64, no. 4, 2002, doi: 10.1086/jop.64.4.1520104.

[6] N. Dian Paramita, "Mewujudkan Kampanye Deliberatif Dalam Sistem Pilkada Serentak 2024 di Indonesia," Jurnal Bawaslu Provinsi Kepulauan Riau, vol. 2, no. 2, 2020, doi: 10.55108/jbk.v2i2.242.

[7] M. Saleh and S. Hadi, "Pengawasan Terintegrasi Terhadap Kampanye Prematur Petahana Dalam Pemilihan Kepala Daerah," Jurnal Adhyasta Pemilu, vol. 3, no. 2, 2021, doi: 10.55108/jap.v3i2.13.

[8] S. Kemp, "Digital 2021: Global Overview Report — DataReportal – Global Digital Insights," Kepios Pte. Ltd., We Are Social Ltd. and Hootsuite Inc.

[9] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," Int J Inf Manage, vol. 35, no. 2, 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.

[10] S. Stieglitz, L. Dang-Xuan, A. Bruns, and C. Neuberger, "Social Media Analytics: An Interdisciplinary Approach and Its Implications for Information Systems," Business & Information Systems Engineering, vol. 6, no. 2, 2014, doi: 10.1007/s12599-014-0315-7.

[11] J. Albors, J. C. Ramos, and J. L. Hervas, "New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source," Int J Inf Manage, vol. 28, no. 3, 2008, doi: 10.1016/j.ijinfomgt.2007.09.006.

[12] J. Tripathy, "Picturing Development: Outdoor Campaign Materials during the 2019 General Election in India," South Asia: Journal of South Asia Studies, vol. 45, no. 4, 2022, doi: 10.1080/00856401.2022.2074034.

[13] P. Kotler and G. Armstrong, Marketing: An Introduction 13th Ed, vol. 59. 2017.

[14] J. E. Spillan, "International Marketing (9th edition), by Michael R. Czinkota and Ilkka A. Ronkainen ," Journal of Teaching in International Business, vol. 21, no. 1, 2010, doi: 10.1080/08975931003644630.

[15] "New Challenges to International Marketing," 2009. doi: 10.1108/s1474-7979(2009)0000020024.

[16] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," Bus Horiz, vol. 53, no. 1, 2010, doi: 10.1016/j.bushor.2009.09.003.

[17] C. P. Hoffmann, C. Lutz, and G. Ranzini, "Privacy cynicism: A new approach to the privacy paradox," Cyberpsychology, vol. 10, no. 4. 2016. doi: 10.5817/CP2016-4-7.

[18] R. Zafarani, M. A. Abbasi, and H. Liu, Social media mining: An introduction, vol. 9781107018853. 2014. doi: 10.1017/CBO9781139088510.
[19] J. Zhao, K. Liu, and L. Xu, "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions," Computational Linguistics, vol. 42, no. 3, 2016, doi: 10.1162/coli_r_00259.
[20] D. Chaffey and P. R. Smith, Digital marketing excellence: Planning, optimizing and integrating online marketing. 2022. doi: 10.4324/9781003009498.
[21] S. Hussain, O. Ameri Sianaki, and N. Ababneh, "A Survey on Conversational Agents/Chatbots Classification and Design Techniques," in Advances in Intelligent Systems and Computing, 2019. doi: 10.1007/978-3-030-15035-8_93.
[22] D. Jurafsky and J. H. Martin, "Constituency Grammars," Speech and Language Processing (3rd ed. draft), 2019.
[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.," Journal of machine Learning research, vol. 3, 2003.
[24] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," Journal of Machine Learning Research, vol. 14, 2013.
[25] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, no. 6755, 1999, doi: 10.1038/44565.
[26] E. Mammoliti, F. Di Stefano, D. Fronzi, A. Mancini, E. S. Malinverni, and A. Tazioli, "A Machine Learning Approach to Extract Rock Mass Discontinuity Orientation and Spacing, from Laser Scanner Point Clouds," Remote Sens (Basel), vol. 14, no. 10, 2022, doi: 10.3390/rs14102365.
[27] S. Sheleha, A. Saputri, H. M. Wicaksana, "Sentiment Analysis with Supervised Topic Modelling on Twitter Data Related to Indonesian Election 2024", doi: 10.1109/IC3INA60834.2023.10285800.