Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 7, No. 2, September 2024, pp. 229 – 240
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Exploratory Data Analysis of Indonesian Presidential Election Candidate Campaign in 2019 on Twitter

[1]Fadlan Bima Hermawan, [2]Taufik Edy Sutanto, [3]Ary Santoso
[1,2,3]Departement of Mathematics, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia
Email: [1]fadlanh32@gmail.com, [2]taufik.sutanto@uinjkt.ac.id, [3]santosoary@uinjkt.ac.id

Article Info

Article history:
Received Jan 22th, 2024
Revised Mar 24th, 2024
Accepted Apr 15th, 2024

Keyword:
Campaign
Exploratory Data Analysis
Politics
Social Media Analysis
Twitter

ABSTRACT

Social media users in Indonesia are growing over time, this has caused many political actors, both individuals, and political parties, to take advantage of this. According to Hootsuite (We are Social) in the Digital 2022 report, Indonesia has 191 million active social media users, so there will be many political actors campaigning on social media. In the literature that discusses similar topics, it is rare to analyze comprehensive exploratory data analysis such as text analysis, hashtags, and social network analysis. From the data analysis conducted in the 2019 Election, the following results were obtained. In text analysis, the narratives used by the two candidates were very different, it was seen that some used other positive and negative narratives. The hashtag analysis found that consistency in campaigning had an influence on electability in the election. In the analysis of social networks, it is found that users who influence social media campaigns, users who appear are not only politicians but also ordinary people who try to support their chosen candidate. The research conducted is expected to be a basic reference in exploring data on social media in other cases, especially in the political field.

Corresponding Author:
Fadlan Bima Hermawan,
Departement of Mathematics,
Univeritas Islam Negeri Syarif Hidayatullah Jakarta,
Jl. Ir H. Juanda No.95, Ciputat, Kec. Ciputat Tim., Kota Tangerang Selatan, Banten 15412.
Email: fadlanh32@gmail.com

DOI: http://dx.doi.org/10.24014/ijaidm.v7i2.26308

1. INTRODUCTION

At this time, the development of technology and information is significantly fast. With the advancement of very sophisticated technology, the growth of social media has finally influenced various activities, even the community can actively contribute to the political event. Social media provides an opportunity for users to create a public opinion that represents the topic being discussed [1]. So that the social media finally become a means and tool for campaigning. Based on reports from several media, one of them is Elsam Multimedia reported there is 116 campaign ad in media social spreaded over in 9 regions in Indonesia [2]. In other words, as reported by pinterpolitik, the use of hashtags on social media is also one way to increase electability during the campaign period.

Indonesia is a country with approximately 276 million people, according to the World Population Review, Indonesia is the fourth country with the largest population in the world. We are currently in the political year leading up to the 2024 Indonesian General Elections, and political contestation will begin soon. As a result, political contestation in Indonesia will involve many elements of society, so to reach more layers of society, new political tools are needed in Indonesia besides conventional campaigns. Social media is an effective tool for political campaigns because it can reach a wide range of voters and supporters [3]. According to Hootsuite (We are Social) in the Digital 2022 report, Indonesia has 191 million active social media users. Therefore, social media is very suitable as a new political tool in Indonesia. According to Ardianto, social media can be used to make the publicity and image of individuals and institutions [4]. So that the image of candidate will be influenced by what they display on social media during campaign

activities [5]. Social media is a place that you can make an opinion freely, so that this platform is a great political tools now.

Brian Heredia et al. in their research explored the effectiveness of social media as a source for polling and predicting election results. They analyzed polls at the national and state levels, the volume of tweets and positive sentiment on tweets are two metrics used to predict election results. The result of this study is that at the national level the volume of tweets is closer to the election results, while at the state level the positive sentiment on the presidential candidate is closer to the election results [6]. A study from Kyoo-Sung Noh, discusses the Big Data-Based Election Campaign Strategy Model in South Korea. In this study, Kyoo-Sung Noh concluded that there are 3 uses of big data in election campaigns. First, increasing communication opportunities for political parties to understand the political attitudes and views of each person more broadly. Second, shaping the agenda in political parties and making political parties able to improve their ability to provide services to the community compared to using traditional surveys, and finally the voter-centric election model using big data can be created and political costs can be reduced to a minimum amount [7]. Wanza et al, conducted an analysis process of SNA (social network analysis) and SA (semantic analysis) which aims to review various techniques in SNA and SA and determine their performance with politics-related topics. After observing various techniques, it was concluded that applying more than one technique can optimize performance without reducing accuracy. However, despite this, there are still areas that future researchers should pay attention to in relation to the influence of politics [8]. Research from Kalamokis et al. aimed to predict the outcome of the 2010 UK election through analysis of Twitter data utilising Linked Open Data. The researchers collected election-related Twitter data using specific hashtags. By identifying entities mentioned in tweets and analysing the volume and distribution patterns of tweets related to political parties, the researchers sought to predict the percentage of votes for the Conservative Party, Liberal Democrats and Labour Party. Through this approach, this research contributes to the understanding of the extent to which election outcomes can be predicted through social media [9]. The research conducted by Al-Gahdir et al. aims to explore the collective behaviour of Arabic social media users, focusing on demographic posting patterns and author profiles. Through analysing metadata from popular local social media forums in Saudi Arabia, this research extracts features to train classifiers in identifying the gender of the author. The results show that the use of 100 features provides the best results, where the 1-NN classification achieves a balanced accuracy of 93.16% in predicting the gender of the author, surpassing the SVM which achieves 87.33%. In addition, this study highlights the importance of avoiding the use of a small k (≤ 50) to avoid over-fitting, as well as confirming the statistical significance of the performance difference between the two classifications [10].

A fascinating part of a country's electoral process is its campaign activities. In this process, valuable information is generally found. In the 2019 Indonesian elections, many issues were associated with the campaign such as politicization of religion, the Indonesian Communist Party (PKI) and even black campaigns. Many previous studies have discussed the role of social media and how to predict election results with it. Therefore, it is interesting to research these issues and find more valuable insights and conduct these experiments in the context of Indonesia. Based on this background, researchers are interested in conducting exploratory data analysis on the 2019 General Election in Indonesia. The researcher will conduct several exploratory analyses to get a real picture of the social media campaign on twitter. So that this research is more focused on finding campaign patterns and strategies on social media by presidential candidate pairs, which is one of the goals of this research that the public can be more educated to use social media, especially during the campaign period. In addition, this research will also conduct social network analysis to find out influential users during the social media campaign.

2. RESEARCH METHOD

This research will discuss the use of exploratory data analysis as a method to explore social media data. The steps will be outlined as clearly as possible so that readers can understand the purpose of this research. The following are the stages that researchers carried out during the research as shown in Figure 1.

2.1. Data and Variable

This research will focus on analyzing social media data, especially twitter. The data used is twitter scrapping data for one year taken from March 1st, 2018 to March 31st, 2019. The data is taken based on 34 provinces in Indonesia using six keywords such as jokowi, jokowilagi, jokowiamin, prabowo, 2019gantipresiden and prabowosandi. The data collected amounted to 1,115,900 data consisting of 6 columns. The following is a display of the initial data scrapping results shown in Table 1.

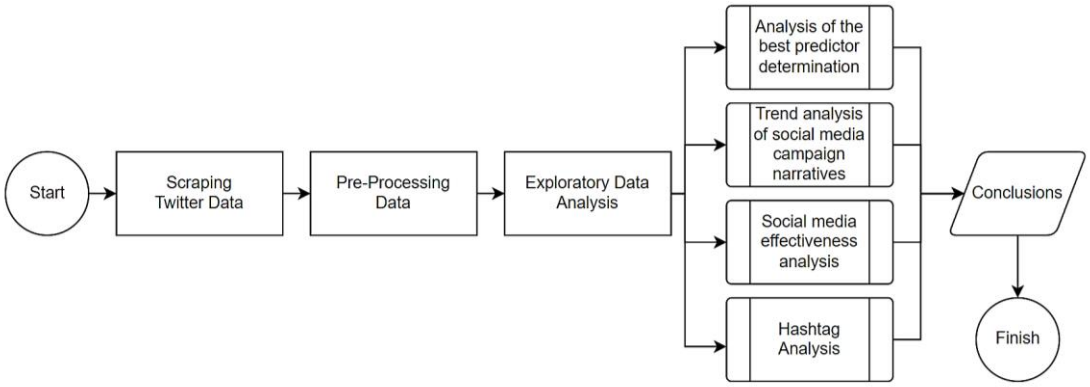


Figure 1. Research Flowchart

Table 1. Example of Data

created_at	username	reply_count	retweet_count	like_count	text
31/03/2019 6:15	gst_kade	0	0	10	@asapardan excelent pak @jokowi, my presiden....
31/03/2019 5:30	dino78073717	0	0	0	@penguntaikata @simkuringporsea @yunartowijaya @prabowo @jokowi #plongaplongo

In addition to data from twitter, this research also uses vote count data for the 2019 presidential election from the Indonesian General Election Commission website. This data is in the form of vote count results for each candidate in each province. Defining variables is important in research, in this study several variables were determined, including Tweet reply count, Tweet retweet count, Tweet like count, Tweet created at, Tweet text dan Tweet username. In addition to the variables already mentioned, this research also pays attention to time variables related to the election campaign process, i.e. soft campaign, hard campaign, and full campaign. Soft campaign is a time variable where the campaign is carried out before the official campaign period is announced by the General Elections Commission in Indonesia called KPU, hard campaign is a time variable where the campaign is carried out during the official campaign period from KPU, and finally the full campaign which is a combination of the two variables above. This research also pays attention to regional aspects where in the analysis of determining the best predictor variables, the data will be divided into three regions, such as Java region, Outside Java region and National to see differences in the tendency of predictor variables in the three selected regions.

2.2. Data Pre-Processing

At this stage, pre-processing will be explained, which is a fairly important stage in text data processing. Social media text data has an unstructured format, so the information cannot be retrieved, so it needs a way to make it more structured. Text preprocessing aims to make the data more structured and reduce noise in the data such as abbreviations and irregular shapes. This stage is very important because the data will be prepared so that further analysis can be carried out, the preprocessing stages carried out in this research are Remove Duplicate, Case Folding, Tokenizing, Remove Stopwords, Fix Slangwords and Stemming [11].

2.3. Pearson Correlation Test

Pearson correlation is a simple correlation that only involves two variables. Pearson correlation produces a correlation coefficient that serves to measure the strength of the linear relationship between two variables [12]. Pearson correlation is used to determine the level or closeness of the relationship between two variables or two object features [13]. The magnitude of the Pearson correlation value can be seen in the following equation:

$$r_{xy} = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

Where r_{xy} is the correlation coefficient, n is the number of samples or observations and x, y are numerical variables

2.4. Exploratory Data Analysis (EDA)

The term exploratory data analysis was first introduced by John Tukey in 1961, exploratory data analysis or EDA is a procedure for analyzing data, a way to make interpretations or interpretations, a plan for obtaining data with the aim of facilitating analysis, and all the equipment and results of mathematical calculations which are used to analyze data [14]. In the process, the data exploration analysis is divided into five parts, namely [15]:

1. Identifying Data Types: Exploratory data analysis can be interpreted as a plan in obtaining data with the aim of facilitating analysis.
2. Viewing Data: A critical point in conducting exploratory analysis of data is how to look at the data. One of the effective ways of making data is visualization, so making visualizations or plots is important in conducting analysis.
3. Outliers: An outlier is data that is significantly different from the rest of the data . Outliers are divided into two there are statistical outliers and there are outliers due to entry errors in the data, and each should be treated very differently otherwise the conclusions drawn from the data may be misleading.
4. Distribution: The frequency of occurrence of a variable value in a sample or population [16]. This distribution can also be depicted in a graph. In this section, the data that has gone through the next three sections will be seen how the data is distributed.
5. Testing: The last part is testing, after going through the previous EDA process then testing can be done, because each test will be different, so it is important to decide which test is best suited to the data.

More generally, EDA covers preprocessing, visualization, to story-telling and reporting, including basic statistical calculations, handling missing values and outliers, dimension reduction, clustering, transformation and data distribution. The exploratory data analysis that will be carried out in this study is divided into 5 stages, such as analysis of the best predictors for real count election data using twitter, analysis of the effectiveness of using social media as a campaign platform, text analysis, hashtag analysis and ended with social network analysis.

3. RESULTS AND ANALYSIS

3.1. Determination of the Best Predictor Variable based on Twitter

The first analysis will be carried out to find the relationship between twitter and election results, especially the 2019 presidential election. In this analysis we use several variables namely Tweet Frequency, Tweet Engagement and Total Twitter User. These three variables will be compared with the vote acquisition data from the Indonesian General Election Commission and then seen the distribution based on Provinces.

1. Predictor Variables Based on Tweet Frequency
   In this analysis, the number of tweets per province is calculated and compared with the vote results found on the General Election Commission website. In addition, in this analysis, the data will be divided into three times to get an overview of the timing of the campaign. This division is based on campaign time, namely soft campaign (SC), hard campaign (HC) and full campaign.

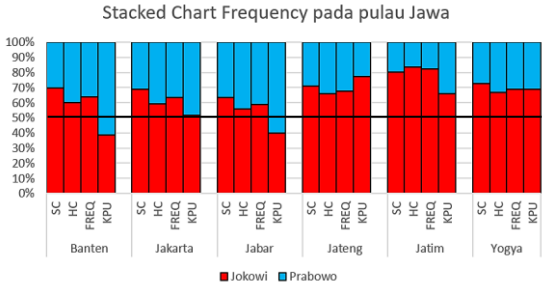


Figure 2. Comparison results of Tweet Frequency with KPU in Java Island

Figure 2 shows an example of the comparison results between the frequency of tweets in each province and the voting results in the same province. In the 2019 presidential election, according to the KPU, Joko Widodo won the votes in 20 provinces while Prabowo Subianto won the votes in the remaining 14 provinces. However, the results of the vote count based on the frequency of tweets stated that all provinces in Indonesia were won by Joko Widodo. So there is an indication that Tweet Frequency is not the best predictor variable for KPU data. To increase confidence in the hypothesis

that has been presented, we conducted a Pearson correlation test and accuracy calculation based on the time variable, then each time variable will be seen based on three regional categories that have been previously mentioned. The highest correlation is on the soft campaign variable at Java Island with a value of 0.612. Furthermore, calculating the accuracy based on the suitability of the results obtained in the frequency analysis compared to the KPU data, the result is in the hard campaign variable of 66%.

2. Predictor Variables Based on Engagement
   Furthermore, we will try to use the number of engagement variables to predict the results of the presidential election. According to twitter business, engagement is the total number of times a person interacts with a tweet. Therefore, in this analysis, data related to engagement will be totaled and then compared with KPU data. The data used as engagement variables are likes, replies and retweets of a post.

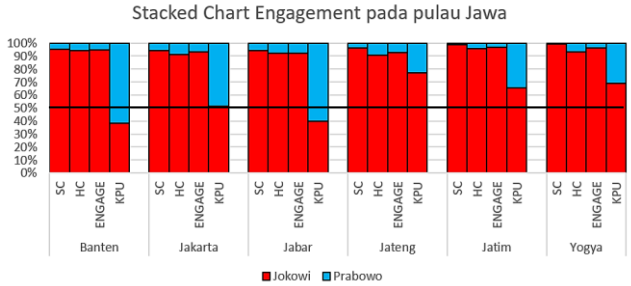


Figure 3. Comparison results of Engagement with KPU in Java Island

Figure 3 shows a very large difference in results between engagement and KPU data, and the difference is even greater than the analysis with the tweet frequency variable. Based on these findings, it can be concluded that during the campaign period, the engagement received by Joko Widodo far exceeded his opponent and indicated that the engagement variable was also not the best predictor of the KPU vote results data. The same thing is also done on this variable to test the hypothesis previously presented, the result is that the best correlation occurs during the soft campaign in Java Island, which is 0.7 while the greatest accuracy is 59% during the soft campaign.

3. Predictor Variables Based on Total Twitter User
   In this analysis, we approached the election process as closely as possible by counting the individuals who voted for a candidate. Twitter users who post about presidential election candidates will be summed up by province and compared to the actual votes. Since we approximated the votes with the total number of users, the soft campaign and hard campaign variables were not used in this analysis.

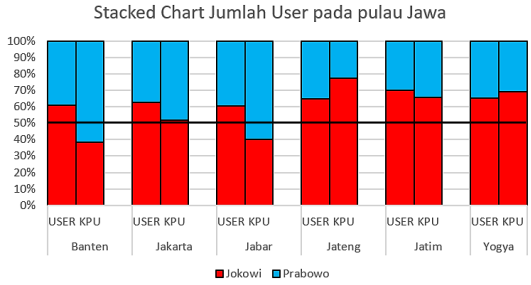


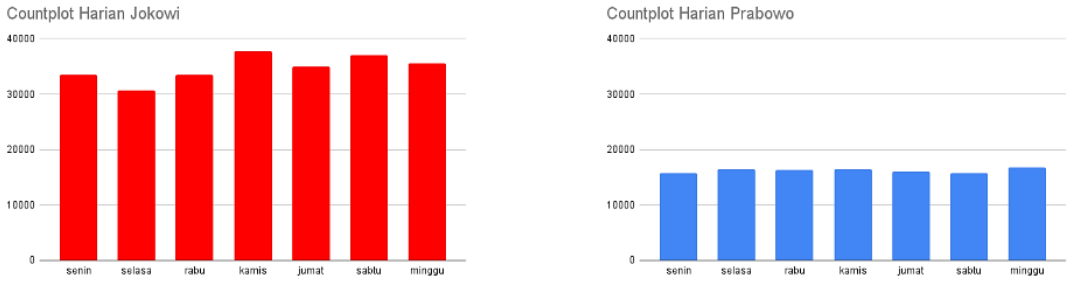
Figure 4. Comparison results of Total User with KPU in Java Island

In general, Figure 4 indicates that the calculation results based on both twitter users and actual votes show a dissimilar pattern. However, the correlation on this variable is the highest compared to other variables, which is 0.716 with a conformity accuracy of 59%.

Based on the results of the analysis using three main variables juxtaposed with three time variables and dividing them into three regions. The result is that the test variable that has the highest correlation with the 2019 presidential election vote is the variable Total users on the full campaign on the island of Java.

3.2. Social Media Effectiveness Analysis

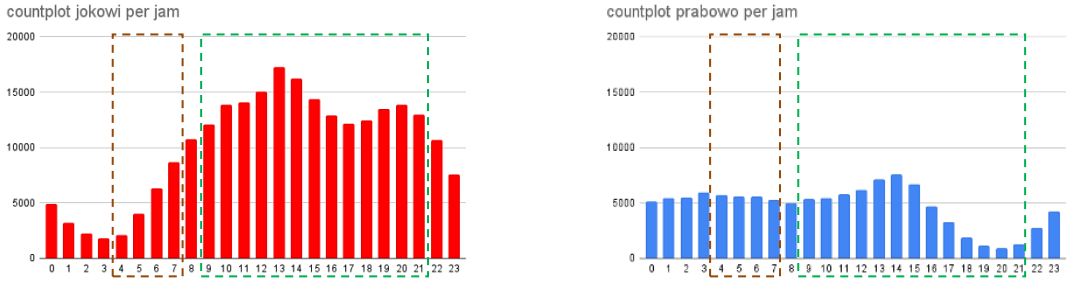
In the second part of the Exploratory Data Analysis series on the 2019 presidential election campaign on Twitter, we will analyze the campaign patterns of each candidate based on the best time to post according to an American marketing company called Hubspot. At this stage the variable used is the time of posting a tweet (created_at). According to an article posted by Hubspot, there is a best time to do marketing or campaigns on social media so that the post gets good engagement or insight. In the article Hubspot provides the best time recommendations based on surveys and research they have done. On Twitter the best day to post is Friday. Then for details 23% of marketers surveyed said that the best posting time was 9 AM - 12 PM, 22% of other marketers said the best time was 12 PM - 3 PM and 21% of marketers said the best time occurred at 3 PM - 6 PM. Based on the article, we also found that the worst time to post on twitter is early morning, especially 6 AM - 9 AM [17]. Hence, if a conclusion is drawn from the article, the best time to post is 9 AM - 6 PM.



(a) **(b)**

Figure 5. Daily Countplot of Campaign Tweets on 2019 **(a)** Jokowi's Countplot; **(b)** Prabowo's Countplot

Figure 5 shows a daily graph of the frequency tweets related to Jokowi and Prabowo. In general, the graph does not show significant differences on a daily basis and is different from Hubspot's recommendations, indicating that the campaign teams from both parties in this case have not used social media effectively. Campaign teams or political figures need to pay attention to this, because if they use strategies in social media campaigns, their engagement may increase and their chances of being elected in political contestation can be higher.



(a) **(b)**

Figure 6. Hourly Countplot of Campaign Tweets on 2019 **(a)** Jokowi's Countplot; **(b)** Prabowo's Countplot

Figure 6 shows the hourly tweet distribution graph in the data, the x-axis in this graph shows the time for 24 hours and the y-axis the frequency of tweets every hour. We have highlighted the figure to make it easier for readers to capture the difference between the two candidates. As mentioned in the previous section, the best time to post on Twitter is 9 AM - 9 PM while the worst time is in the early morning. The result is that posts about Joko Widodo are more in line with the recommendations given. On the other hand, posts about Prabowo Subianto tended to be posted during midnight to early morning and did not seem to match Hubstpot's recommendations. Based on the data and analysis conducted, it can be concluded that if it is assumed that the user who posted about Joko Widodo is his social media campaign team, then it shows that Joko Widodo's social media campaign team uses the strategy of posting tweets at the most appropriate time and using social media more optimally than his opponent.

3.3. Narrative Trend in Social Media Campaigns

In political contestation, there are always narratives that want to be built and conveyed to the public by candidates for public officials, so it is necessary to analyze the narratives carried by the two candidates for

the 2019 presidential election. To see the narratives built, researchers conducted text analysis on the data. Before being analyzed, the data was separated into three variables based on time, namely soft campaign, hard campaign and full campaign. The definitions of the three variables have been explained in the previous analysis.

1. Narratives Analysis
   In this analysis, researchers will try to see and understand the narratives that appear in the campaigns of the two candidates for the 2019 presidential election. The tool that will be used to get the desired analysis results beforehand is word cloud. WordCloud is a Cytoscape extension designed to offer a graphical overview of annotations, facilitating the analysis and interpretation of networks [18], words with a larger size that appear in word cloud indicate that the word appears frequently in the text.

(a) **(b)**

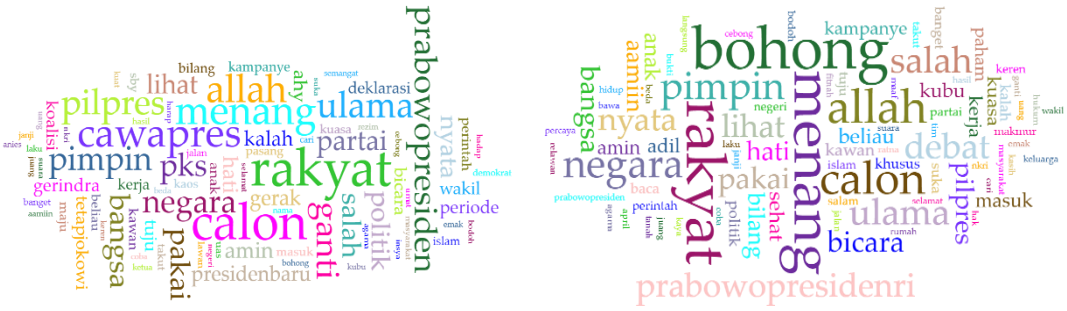
Figure 7. Word cloud Narratives Full Campaign (a) Jokowi's Wordcloud; (b) Prabowo's Wordcloud

Figure 7 is Joko Widodo's wordcloud during the full campaign. In social media analysis, many literatures say that social media is not effective in villages because internet access is still difficult in these areas and social media has not reached these areas. Because of this, villages are not a priority in conducting social media campaigns. However, Figure 7 (a) which displays a list of words in social media texts related to Joko Widodo shows different results. The results show that village-related issues often appear such as jokowi membangun desa, desa, kabupaten, camat, etc. This indicates that village issues were crucial and important issues to be discussed during Jokowi's campaign, and that village involvement was important here. Figure 7 (b) shows the wordcloud of full campaign variables related to Prabowo Subianto, the set of words that appear most are related to religious issues such as ulama, islam, hati, Allah and so on. This is in line with what happened in the 2019 election, that Prabowo Subianto's campaign team did use religious issues as the main weapon.

(a) **(b)**

Figure 8. Jokowi's Word cloud Narratives (a) Soft Campaign; (b) Hard Campaign

It can be seen in Figure 8 that the narrative used looks consistent because in this case village-related issues remain a topic of conversation related to Joko Widodo, this indicates that the issue of villages is used by Joko Widodo's campaign team, especially on social media to improve Jokowi's reputation in the 2019 presidential election.



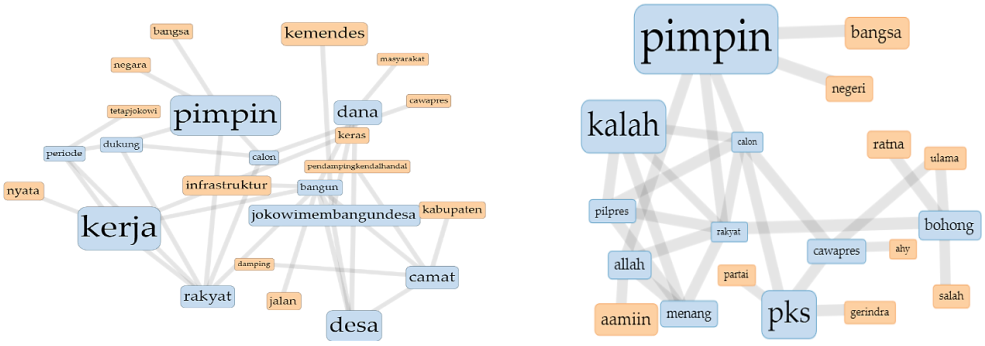
(a) **(b)**

Figure 9. Prabowo's Word cloud Narratives (a) Soft Campaign; (b) Hard Campaign

It can be seen in Figure 9 that the narrative used by Prabowo seems inconsistent, this is because there are quite a lot of large differences in word frequency in both wordclouds. If we look more clearly, the narratives used during the soft campaign tend to be more positive because they try to increase the personal branding of Prabowo by building narratives such as Prabowopresiden and presiden baru. However, during the hard campaign the narratives used tended to be more negative such as bohong, cebong, bodoh dan takut, this shows the difference in narratives during soft and hard campaigns. This shows that there are indications that during the hard campaign Prabowo's social media campaign team used attacks to bring down political opponents, but this needs to be seen more deeply in the meaning and context contained in the word so that the above conclusions can be scientifically proven.

2. Context Analysis

Next, the researcher will try to get another context from the narrative that has been obtained from the previous analysis to get the issue. The tool to visualize this other context is wordlink. WordBridge or Wordlink is an innovative graph-based visualization method aimed at illustrating connections among entities within textual corpora. This technique adopts a node-link visualization approach wherein both nodes and links are represented as tag clouds [19]. Wordlink is one of the text visualization methods that serves to see related words from each text that appears in the data.

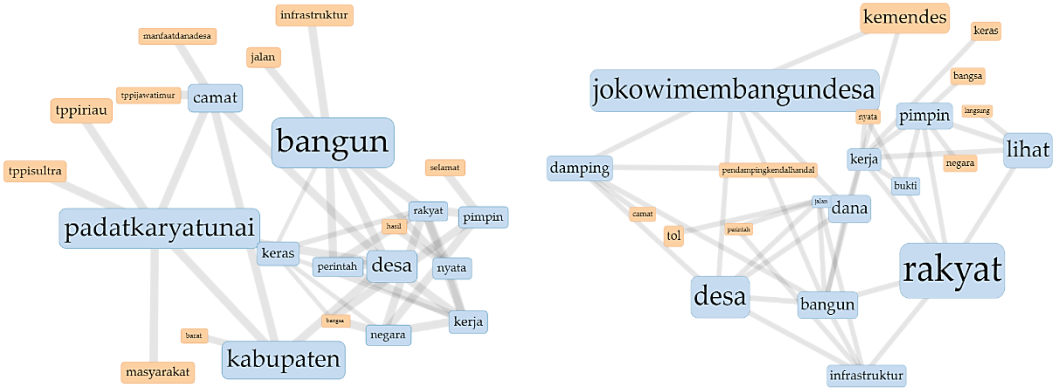


(a) **(b)**

Figure 10. Wordlink Narratives Full Campaign (a) Jokowi's Wordlink; (b) Prabowo's Wordlink

In wordlink, the different color labels of words indicate the different status of the word, a blue labeled word is a primary word while an orange labeled word is a secondary word or a word that is often paired with a primary word. In addition, the difference in line thickness also has its own meaning, if the connecting line between words is thick then the words often appear in the same sentence, otherwise if the connecting line is thin then the connected words appear less in the same sentence. Based on the visualization shown in Figure 10, it is found that there is a difference in the number of words that appear on both Wordlinks. In the wordlink of Joko Widodo, more words are displayed, which means that the phrases in the text related to Joko Widodo are more varied. This is supported by thin lines connecting words, which means that the words that appear are not often in one sentence, so there are indications that the narrative built by Jokowi's social media campaign team varies. A different condition was found in Prabowo Subianto's wordlink, which tended to show few words and there was a very dominant word, "pimpin". From this, it can be indicated that the

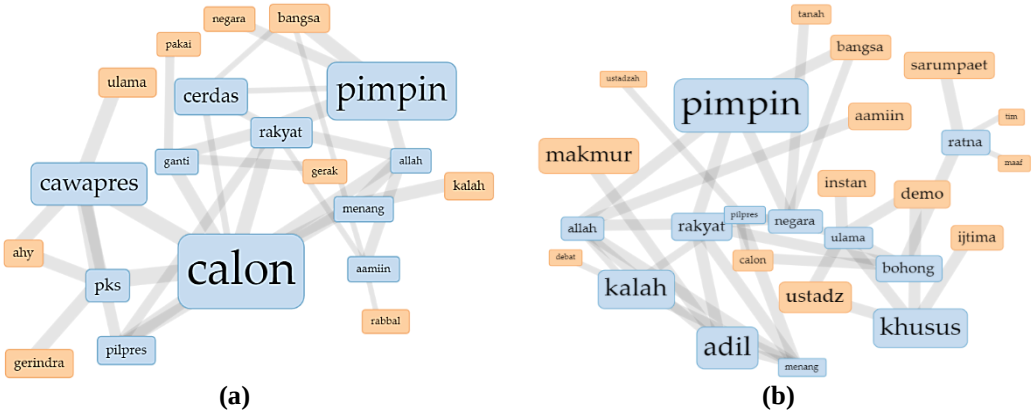
narrative that was trying to be built in the campaign about Prabowo Subianto was related to leadership. Another interesting point is that the connecting lines between words that tend to be thick indicate that this campaign does not use varied narratives and tends to only carry one theme in its social media campaign process.



(a) **(b)**

Figure 11. Jokowi's Wordlink Narratives (a) Soft Campaign; (b) Hard Campaign

Both of Joko Widodo's wordlinks show that there is no significant change between words in the soft campaign and the hard campaign, this can be seen in Figure 11. It can be seen that the issue of villages is an issue that is consistently discussed. It can be seen that the lines connecting the words remain relatively thin, but the same issue is still built. In this case, the village issue is the main issue, but it seems that the narrative on this village issue is varied so as to make the subconscious of potential voters think that Jokowi is focused on developing villages and disadvantaged areas.



(a) **(b)**

Figure 12. Prabowo's Wordlink Narratives (a) Soft Campaign; (b) Hard Campaign

In general, Figure 12 shows that the issues discussed during the soft campaign tend to be different from the hard campaign. During the soft campaign, the narratives used were more inclined towards emphasizing Prabowo Subianto. However, during the hard campaign, the narrative that was built led to identity politics and the use of negative issues to attack political opponents.

Based on the analysis that has been done in this section, several conclusions can be drawn. First, by looking at the narrative using wordcloud, it seems that the consistency of the use of narrative when campaigning on social media, especially twitter, is important for victory in politics, this is concluded because it sees the difference in the consistency of the narrative carried between Jokowi and Prabowo on twitter social media. Furthermore, in looking at the context using wordlinks, it is clear that what has been concluded above is further strengthened by the words and context that appear in the wordlinks. In addition to consistency in the narrative, it seems that it would be much better if there were variations in building the narrative. Finally, strategic issues and narratives developed based on track records during office are indicated to be able to influence the outcome of political contestation.

3.4. Hashtag Analysis

In social media conversations, hashtags function as indicators of topics, markers of interpersonal relationships, and organizers of text, thereby improving the accessibility of social content and enriching the communicative potential of social media interactions [20]. Based on this definition, using hashtags to analyse topic distribution on social is very relevant, then we analyzed the use of hashtags to find out how much influence hashtags have on social media campaigns.

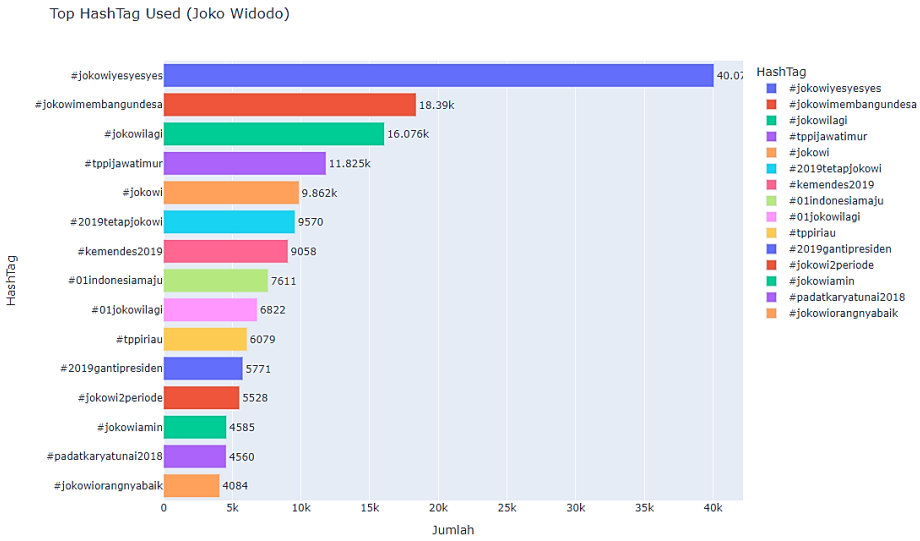


Figure 13. Top 15 Hashtag Related to Joko Widodo's Tweet

Previously in wordcloud and wordlink we saw that village-related issues were often discussed, this seems to be in line with the results of the hashtag collection that has been carried out seen in Figure 13 that quite a lot of hashtags raise this issue, which means that the suspicion that village issues are used for campaigning is getting stronger. It can be seen that several hashtags functioned to emphasize Joko Widodo such as #Jokowi and #Jokowiorangnyabaik and the hashtags shown also consistently discussed Joko Widodo. In addition, the narrative that is trying to be built seems to be positive as seen from the hashtags that appear such as #indonesiamaju and #01optimisindonesiamaju. Because of these things, it seems that the use of hashtags in Joko Widodo's social media campaign is good.

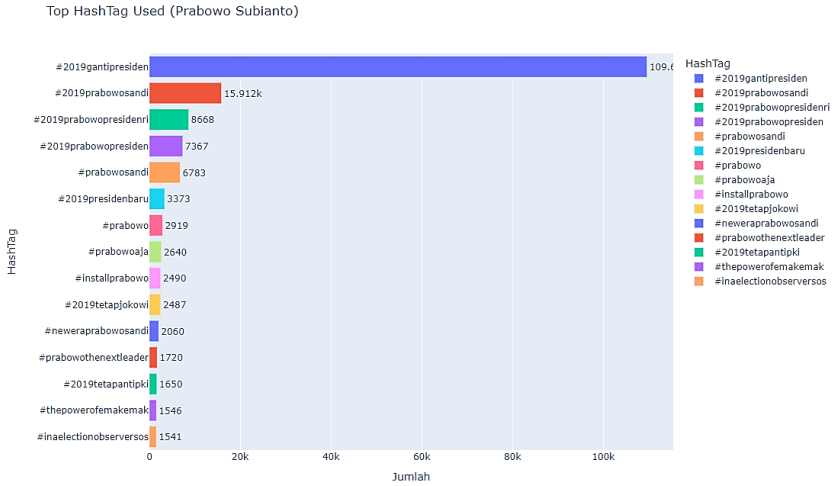


Figure 14. Top 15 Hashtag Related to Prabowo Subianto's Tweet

On the other hand, there is a hashtag used in tweets related to Prabowo Subianto in Figure 14 that is very focused on #2019gantipresiden which when compared to other hashtags its use is very far, which can be concluded that every post for Prabowo Subianto uses this hashtag. In addition, the hashtags that appeared also seemed to only focus on the personalities of the presidential candidate, which is natural because

Prabowo is the opposition and seems unable to prove his achievements. The interesting thing here is that hashtags related to Joko Widodo appeared, such as #2019tetapjokowi, which is different from the hashtags that appeared in tweets related to Joko Widodo, where none of the hashtags mentioned Prabowo Subianto. Finally, a negative hashtag appeared, #2019tetapantipki, which was indicated to be used to attack political opponents.

Based on the analysis above, the conclusion drawn is again related to the consistency of the use of narratives and campaign objectives. It can also be seen in both images that there are negative hashtags in tweets related to Prabowo Subianto, so it seems that it will have a bad impact on candidates if they use negative narratives. Meanwhile, in the hashtags related to Jokowi, the narrative tended to be positive and had a good impact on his electability based on the results of the 2019 presidential election.

4. CONCLUSION

Exploratory Data Analysis conducted in this research provides satisfactory results, this research is expected to be able to find strategies and patterns of candidates in social media. Based on the analysis conducted, patterns of social media campaigns, especially on Twitter, can be found. Campaign patterns such as consistency in building narratives, posting strategies at the best time, track record-based narratives, positive campaigns and the use of strategic issues seem to affect the final results of the presidential election contestation. These patterns can be a recommendation to political actors in developing campaign plans on Twitter to get more effective results. However, this research faces challenges in finding predictors of presidential election votes based on the variables in the Twitter data. In the future, this research can be developed using machine learning models so that several hypotheses formed from this research can be validated.

REFERENCES

[1] N. Anstead and B. O'Loughlin, "Social media analysis and public opinion: The 2010 UK general election," *J. Comput. Commun.*, vol. 20, no. 2, pp. 204–220, 2015, doi: 10.1111/jcc4.12102.

[2] Elsam Multimedia, "Pemanfaatan Media Sosial dalam Iklan Kampanye Politik Pemilu 2019 (Bagian I)," 2020. Pemanfaatan Media Sosial dalam Iklan Kampanye Politik Pemilu 2019 (Bagian I) (accessed Oct. 10, 2022).

[3] L. A. Abdillah, "Social Media As Political Party Campaign in Indonesia," no. 12, pp. 1–10, 2014.

[4] E. A. and B. Q-Anees, *Filsafat Ilmu Komunikasi*, 4th ed. Bandung: Simbiosa Rekatama Media, 2014.

[5] I. H. Harahap, "Kampanye Pilpres 2019 Melalui Media Sosial Dan Pengaruhnya Terhadap Demokrasi Indonesia," *Komunikologi J. Ilm. Ilmu Komun.*, vol. 17, no. 1, pp. 1–11, 2020.

[6] B. Heredia, J. D. Prusa, and T. M. Khoshgoftaar, "Social media for polling and predicting United States election outcome," *Soc. Netw. Anal. Min.*, vol. 8, no. 1, p. 0, 2018, doi: 10.1007/s13278-018-0525-y.

[7] K.-S. Noh, "A Exploratory Study on Big-data based Election Campaign Strategy Model in South Korea," *J. Digit. Policy Manag.*, vol. 11, no. 12, pp. 113–120, 2013, doi: 10.14400/jdpm.2013.11.12.113.

[8] I. Wanza, I. Kamuti, D. Gichohi, and K. Gikunda, "The impact of Twitter on political influence on the choice of a running mate: Social Network Analysis and Semantic Analysis -- A Review," 2022, [Online]. Available: https://arxiv.org/abs/2208.00479.

[9] E. Kalampokis, A. Karamanou, E. Tambouris, and K. Tarabanis, "On predicting election results using twitter and linked open data: The case of the UK 2010 election," *J. Univers. Comput. Sci.*, vol. 23, no. 3, pp. 280–303, 2017.

[10] A. I. Al-Ghadir and A. M. Azmi, "A Study of Arabic Social Media Users—Posting Behavior and Author's Gender Prediction," *Cognit. Comput.*, vol. 11, no. 1, pp. 71–86, 2019, doi: 10.1007/s12559-018-9592-7.

[11] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 874, no. 1, doi: 10.1088/1757-899X/874/1/012017.

[12] M. J. de Smith, *Statistical analysis handbook. A comprehensive handbook of statistical concepts, techniques and software tools*, vol. 37, no. 3. 2018.

[13] P. K. Sen and W. J. Conover, *Practical Nonparametric Statistics.*, vol. 67, no. 337. 1972.

[14] J. W. Tukey, *Exploratory Data Analysis by John W. Tukey*, 1st ed. Pearson, 1977.

[15] V. Cox, *Translating Statistics to Make Decisions A Guide for the Non-Statistician*. 2017.

[16] F. E. Grubbs, "Procedures for Detecting Outlying Observations in Samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969, doi: 10.1080/00401706.1969.10490657.

[17] F. Needle, "Global Social Media Trends Report," New York, 2023.

[18] L. Oesper, D. Merico, R. Isserlin, and G. D. Bader, "WordCloud: A Cytoscape plugin to create a visual semantic summary of networks," *Source Code Biol. Med.*, vol. 6, pp. 2–5, 2011, doi: 10.1186/1751-0473-6-7.

[19] K. T. Kim, S. Ko, N. Elmqvist, and D. S. Ebert, "WordBridge: Using composite tag clouds in node-link diagrams for visualizing content and relations in text corpora," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, pp. 1–8, 2011, doi: 10.1109/HICSS.2011.499.

[20] M. Zappavigna, "Searchable talk: the linguistic functions of hashtags," *Soc. Semiot.*, vol. 25, no. 3, pp. 274–291, 2015, doi: 10.1080/10350330.2014.996948.

BIBLIOGRAPHY OF AUTHORS

Fadlan Bima Hermawan is a graduate of the mathematics study program at Syarif Hidayatullah State Islamic University Jakarta. He currently works as a data analyst in a consulting company. His areas of interest are artificial intelligence, data mining, political science and data government.

Taufik Edy Sutanto is a senior lecturer at the mathematics department of Syarif Hidayatullah State Islamic University Jakarta. He took his undergraduate education at the University of Indonesia with a major in mathematics, S2 at the University of South Wales with a major in statistical computing. He has also completed his Ph.D program in data mining and big data. He actively researches and develops models and technologies related to data mining and big data.

Ary Santoso is a lecturer in the mathematics department at Syarif Hidayatullah State Islamic University Jakarta. He took his undergraduate and postgraduate education at Bogor Agricultural University majoring in statistics. He continued his studies at Padjajaran University majoring in Political Science. His research focuses on statistics, political science and data science.