Sentiment Analysis of YouTube Videos in the 2024 Indonesian Presidential Election

Zaidan Yahya[1], Tomoyosi Akiba[1], Yasutomo Kimura[2]
Yuki Mikiya[3], Kota Mori[4], Mitsuo Yoshida[5], Yuko Kasuya[3]
[1]Toyohashi University of Technology, [2]Otaru University of Commerce
[3]Keio University, [4]Japan Data Science Consortium Co. Ltd., [5]University of Tsukuba
{zaidan.yahya.io,akiba.tomoyoshi.tk}@tut.jp

Abstract

Social media platforms like YouTube have become powerful tools for shaping public opinion during elections in recent years. This study examines the sentiments in the YouTube videos concerning three presidential candidates in the 2024 Indonesian election. We first classify the videos into three categories: candidate's official channel, public news, and third-party-created sources. Next, we perform sentiment analysis on each video and calculate a metric, the Sentiment Impact Score (SIS), to quantify the overall sentiment dynamics. Our findings reveal a significant shift in public sentiment, ultimately favoring the elected candidate, especially among the third-party created videos.

1 Introduction

Social media platforms like YouTube, Facebook, and TikTok have become powerful tools for shaping public perceptions of candidates during election campaigns. The Indonesian presidential election, held on February 14, 2024, serves as a poignant example of the power of social media platforms. This study focuses on YouTube videos about presidential candidates during the campaign period.

Several studies have analyzed YouTube videos in the context of the 2024 Indonesian elections. Armayudi et al. [1] focus on analyzing public sentiment on YouTube regarding the presidential candidates' debate. Ma'aly et al. [2] conducted a comprehensive analysis of sentiment surrounding the 2024 Indonesian presidential election.

While above studies primarily focus on the comments section of the videos, this study takes a novel approach by analyzing the videos themselves, focusing on content related to the three presidential candidates, namely, Anies Baswedan, Prabowo Subianto, and Ganjar Pranowo, over several months leading up to polling day.

An important consideration when studying such platforms is that some channels may disseminate videos to manipulate voter perception, including the use of misinformation. To address this, we categorized video sources into three distinct groups: candidate's official channel (**official**), public news (**news**), and third-party-created sources (**third-party**). This categorization enables us to examine differences across these sources. Furthermore, third-party videos are subdivided into two categories: news-like videos and other types of videos. We utilized the second phase of IndoBERT$_{\mathrm{BASE}}$ as our base model for video classifier. IndoBERT [3] is a language model for Indonesian based on BERT [4]. It was trained on a dataset called Indo4B, which contains approximately 4 billion words [3].

Indonesia's election campaign spans several months, and candidate information evolves over time [5]. To track these changes, we employ a large language model to determine the sentiment of each video and calculate the Sentiment Impact Score (SIS) [5] to understand the overall sentiment for each candidate throughout the election period.

2 Sentiment Impact Score (SIS)

Sentiment Impact Score (SIS) [5] is a quantification index that considers the sentiment and article frequency associated with each candidate in an election. After classifying article sentiment as positive, negative, or neutral, SIS is calculated as follows:

$$SIS = \left(\frac{\omega - \psi}{\phi}\right) \times \log(\phi) \tag{1}$$

where ω, ψ, and ϕ are the number of positive articles, negative articles, and all articles excluding neutrals, respectively. More positive articles lead to positive sentiment

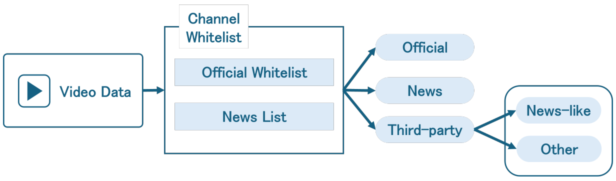

Figure 1 Video Classification Flow

scores, and more negative articles will lead to more negative sentiment scores. Neutral articles are excluded from the calculation because they are considered to have little influence on shaping public opinion. Lastly, greater media coverage will have a more significant effect on the score.

3 Data

YouTube videos were searched using the names of the three presidential candidates, along with the Indonesian terms for **presidential election** and **election**. The videos were collected between late November 2023 and early June 2024. We conducted the classification and sentiment analysis of the video transcript generated using a general-purpose speech recognition model Whisper [6]. A total of 72,844 videos were initially collected; however, due to constraints in Whisper API usage transcripts are available for 36,365 videos. Among these, 102 videos have blank transcripts. This study targets these 36,365 videos.

4 Analysis Method

4.1 Video Classification Method

We aim to clarify the information sources by classifying videos into three main categories: **official**, **news**, and **third-party**. Furthermore, we categorized **third-party** videos into **news-like** and **other** categories. Figure 1 shows these general flows.

Our classification approach uses channel names to create a whitelist of official and news channels. Videos from these channels are not included in the whitelist and are classified as third-party. The **official** category includes videos from presidential and vice-presidential candidates' official campaign channels. The **news** category consists of videos from authorized news media. The **third-party** category consists of those uploaded by individuals or organizations not classified as official or news.

We further categorized third-party videos into two subcategories: news-like and other types. A news-like video is defined as a news clip from third-party channels or a partial recording of a news broadcast. To achieve this, we are developing a video classifier model that uses transcripts to identify news-like videos rather than relying on channel names. However, the model was trained by categorizing all videos into news or third-party channels, a binary classification. This dual-task approach may impact performance, as distinguishing between the news or third-party channels and further splitting third-party videos into news-like and other type are two different tasks which present distinct challenges.

To initiate classification process, we randomly selected 1,000 videos and used this data to create a whitelist of official and news channels. Our whitelist consists of 7 official channels and 80 news channels.

Table 1 Video Classifier Training Data

Label	Videos	Channels
News	2153	76
Third-party	6847	3982
Total	10,000	4,058

Table 2 Video Classifier Validation Data

Label	Videos	Channels
News	244	53
Third-party	756	638
Total	1,000	691

4.2 Experiment Setup

The second phase of IndoBERT$_{\text{BASE}}$ is used as our base model with the following steps. First, we initialized a pseudo-label for each video based on the whitelist we created previously. We then used these data to create the training, validation, and evaluation datasets. For training and validation, we randomly selected 10,000 video transcripts, dividing the dataset into a 9:1 ratio. This resulted in 9,000 transcripts for training and 1,000 transcripts for validation. Tables 1 and 2 show the detailed composition of the training and validation datasets, respectively.

We randomly selected an additional set of 1,000 videos as evaluation data, distinct from those used to create the whitelist. This dataset was used to evaluate the model's performance in categorizing videos into news and third-party categories, referred to as **Test A**. Subsequently, we randomly selected 220 videos categorized as third-party

Table 3 Classification Evaluation Dataset (Test A)

Label	Videos	Channels
News	220	56
Third-party	774	656
Total	994	712

Table 4 Classification Evaluation Dataset (Test B)

Label	Videos	Channels
News-like	18	16
Others	202	185
Total	220	201

in Test A to identify any news-like videos. This extended evaluation dataset was labeled as **Test B**. Tables 3 and 4 show Test A and Test B composition in detail, respectively. Test A allows us to evaluate the video classifier model's performance in categorizing videos into news or third-party videos. At the same time, Test B helps us evaluate the model's ability to detect news-like videos by third parties.

4.3 Sentiment Analysis Method

We selected the Indonesian RoBERTa base sentiment classifier [7] as our model for sentiment classification due to its strong performance in the SmSA dataset [3]. To ensure the model's suitability for our current dataset, we evaluated its accuracy and effectiveness using a manually labeled evaluation dataset.

For the evaluation, we randomly selected 100 YouTube video transcripts classified as third-party. These were manually labeled into five categories: positive, negative, negative/positive, neutral, and undetermined. The negative/positive category indicates a shift in sentiment within the same video. Table 5 provides a detailed breakdown of the evaluation dataset. During the evaluation process, video transcripts labeled negative/positive were evaluated with two conditions. It is evaluated as positive or negative one at a time. After excluding undetermined data, 88 labeled transcripts were used for the final evaluation.

After the evaluation, we applied the sentiment model to analyze the sentiment of all video transcripts and calculated the SIS over time, based on the cumulative number of videos published weekly. This approach enabled us to track changes in sentiment trends during the campaign period.

Table 5 Sentiment Evaluation Dataset

Label	Videos
positive	34
negative	33
negative/positive	12
neutral	9
undetermined	12

Table 6 Video Classification Evaluation Results

Dataset	Accuracy	Precision	Recall	F1 Score
Test A	0.85	0.67	0.62	0.64
Test B	0.91	0.4	0.22	0.29

5 Analysis Results

5.1 Video Classification Evaluation

Video classification model evaluation result is shown in Table 6.

For Test A, which closely aligns with the whitelist results, the model achieved an accuracy of 0.85, indicating strong overall performance in correctly classifying videos. Based on precision and recall, the model demonstrated a reasonable ability to differentiate between news videos and third-party videos using transcripts in our current dataset. However, the model showed lower performance on Test B, particularly on precision and recall. While the accuracy remained high, the low precision and recall indicate that the model struggles to accurately classify news-like videos within the third-party category.

While we trained our model to find news-like videos within third-party category by training the model to classify videos as either news or third-party, the model struggled to find news-like videos. As news-like videos are an underrepresented class, we might consider using video metadata such as descriptions or tags as features to improve the model performance to find news-like videos in the future.

5.2 Sentiment Analysis

5.2.1 Model Evaluation

Table 7 presents the sentiment evaluation results under two different conditions: when negative/positive sentiments are considered positive and when they are consid-

Table 7 Sentiment Evaluation Results for Each Condition

Condition	Accuracy	F1 Score
negative/positive as positive	0.7614	0.7612
negative/positive as negative	0.8295	0.8084

Table 8 Sentiment Analysis Result

Label	Videos	Percentage (%)
positive	13409	36.87
negative	15635	42.99
neutral	7219	19.85
blank	102	0.28
Total	36365	100%

ered negative. The table compares the accuracy and F1 scores for each condition. The evaluation results highlight the model's varying performance under these different approaches to classifying sentiment. Notably, Table 7 demonstrates that the model performs well in identifying sentiments for both scenarios.

Evaluating negative/positive sentiments as negative may increase the model's evaluation performance as shown in Table 7. However, when considering the overall sentiment of the labeled video transcripts, the sentiment negative/positive labeled videos tend to lean more toward positive than negative.

5.2.2 Sentiments for All Videos

The sentiment analysis results, presented in Table 8, illustrate the distribution of sentiment across a total of 36,365 videos. The majority of the videos are classified as negative (42.99%), followed by positive (36.87%) and neutral (19.85%) videos. While there is a small portion (0.28%) of transcripts which are blank as mentioned previously.

5.2.3 Sentiment Impact Score Result

In this analysis, we focused on third-party videos for the SIS. Figure 2 illustrates the SIS based on the cumulative number of published videos over time for third-party videos. Additional results are available in the Appendix A.

The top graph in Figure 2 displays the SIS for each candidate, while the bottom graph shows the cumulative number of published videos related to each candidate. The vertical dashed red line represents the election day, and the vertical dotted black line indicates the day when the results

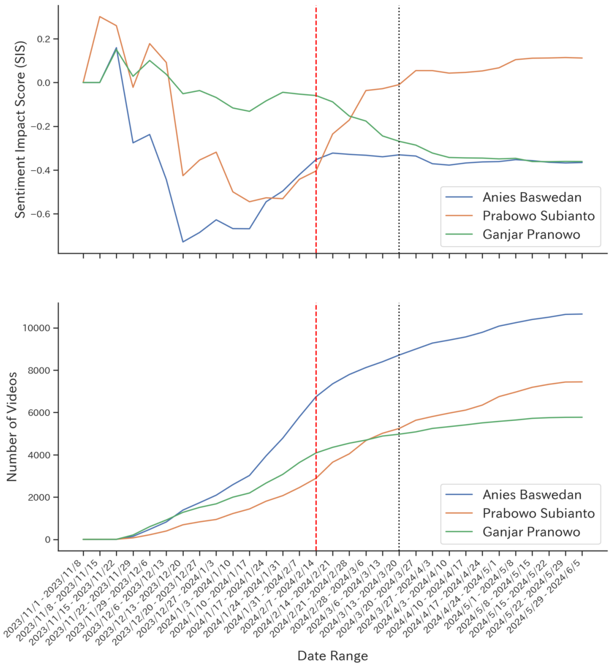

Figure 2 SIS Based on The Cumulative Number of Published Videos (Third-Party)

were announced officially.

As shown in Figure 2, changes in cumulative SIS over time reveal distinct trend. Approximately three weeks before election day, favorable sentiment towards Anies Baswedan and Prabowo Subianto increased. After the election, favorable sentiment towards Prabowo Subianto continued to rise, while negative sentiment towards the other candidates intensified or stagnated overall. These SIS results highlight the shift in public sentiment favoring the elected presidential candidate, Prabowo Subianto, around the time of the polling date.

6 Conclusion

This study explored the sentiments embedded in the YouTube videos regarding three presidential candidates in the 2024 Indonesian election. Among the three video categories, we focused on the third-party videos' sentiment trends. From the SIS results, we revealed a shift in sentiment favoring the winning presidential candidate, Prabowo Subianto, around the time of the polling date and thereafter.

These findings provide a foundation to interpret public sentiment during the election and enable more detailed analyses. However, this method is sensitive to several factors, including whitelist coverage, video classifier performance, and sentiment analysis accuracy. Future research should address these issues.

Acknowledgements

This work was supported by JSPS Topic-Setting Program to Advance Cutting-Edge Humanities and Social Sciences Research Grant Number JPJS00123811919.

References

[1] Andi Armayudi Syam and Andi Ahmad Malikul Afdal. Examining social media public sentiment: Youtube response to 2024 indonesian presidential debate. **KYBERNOLOGY : Journal of Government Studies**, Vol. 4, No. 1, pp. 15–28, 2024.

[2] Ahmad Nahid Ma'aly, Dita Pramesti, Ariadani Dwi Fathurahman, and Hanif Fakhrurroja. Exploring sentiment analysis for the indonesian presidential election through online reviews using multi-label classification with a deep learning algorithm. **Information**, Vol. 15, No. 11, November 2024.

[3] Bryan Wilie, Karissa Vincentio, Genta Indra Winata, Samuel Cahyawijaya, X. Li, Zhi Yuan Lim, S. Soleman, R. Mahendra, Pascale Fung, Syafri Bahar, and A. Purwarianti. Indonlu: Benchmark and resources for evaluating indonesian natural language understanding. In **Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing**, 2020.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[5] Missie Chercheur and Malkenzie Bovafiz. Leveraging ai and sentiment analysis for forecasting election outcomes in mauritius, 2024.

[6] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision, 2022.

[7] Wilson Wongso. indonesian-roberta-base-sentiment-classifier, 2023.

A Appendix

A.1 SIS for All Videos

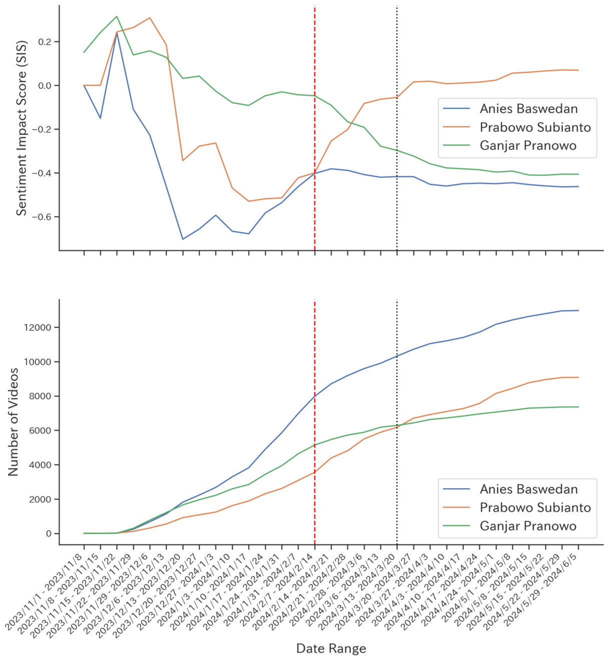

Figure 3 SIS Based on The Cumulative Number of Published Videos

A.2 Temporal SIS

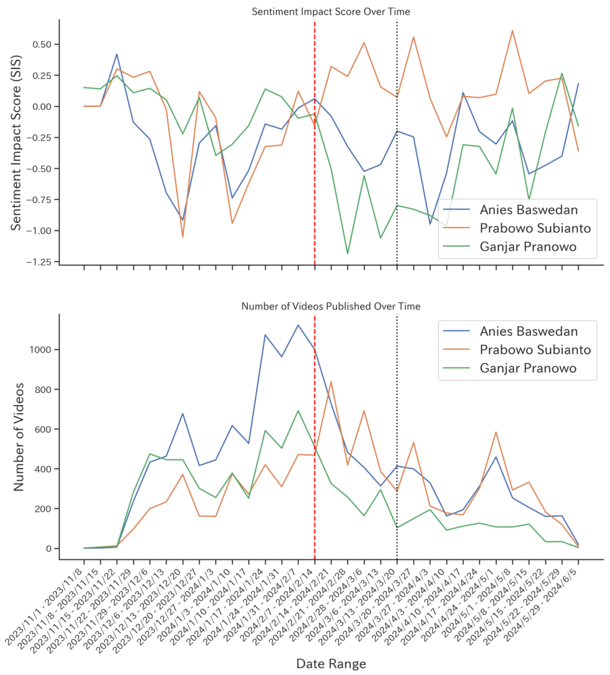

Figure 4 SIS Based on Temporal Number of Published Videos

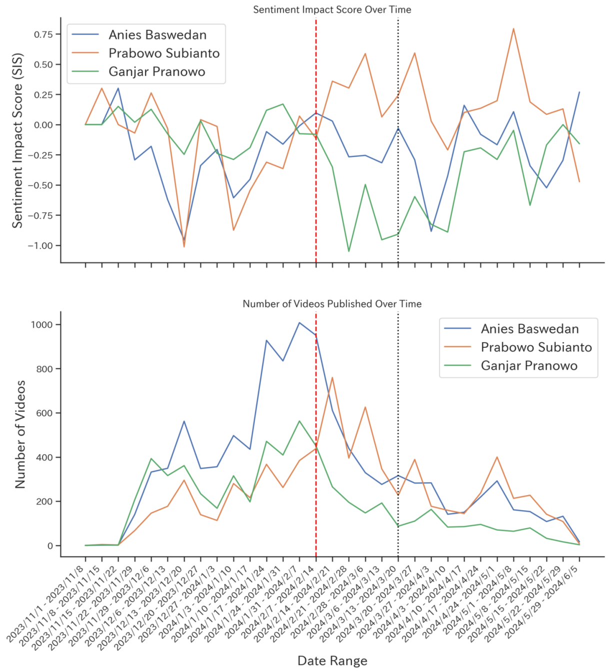

Figure 5 SIS Based on Temporal Number of Published Videos (Third-Party)