Time Series Sentiment Analysis of YouTube Videos in the 2024 Indonesian Presidential Election

Zaidan Yahya[1], Tomoyoshi 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

2025/09/20

Background and Objectives

- Social media platforms have become powerful tools for shaping public perceptions of candidates during election campaigns
- Some studies focus on the comments section of the YouTube videos concerning the 2024 Indonesian election

- This study examines YouTube videos itself, concerning three presidential candidates in the 2024 Indonesian election
- Two analyses were conducted using video transcripts: **Video Classification** and **Sentiment Analysis**

The 2024 Indonesian Presidential Election

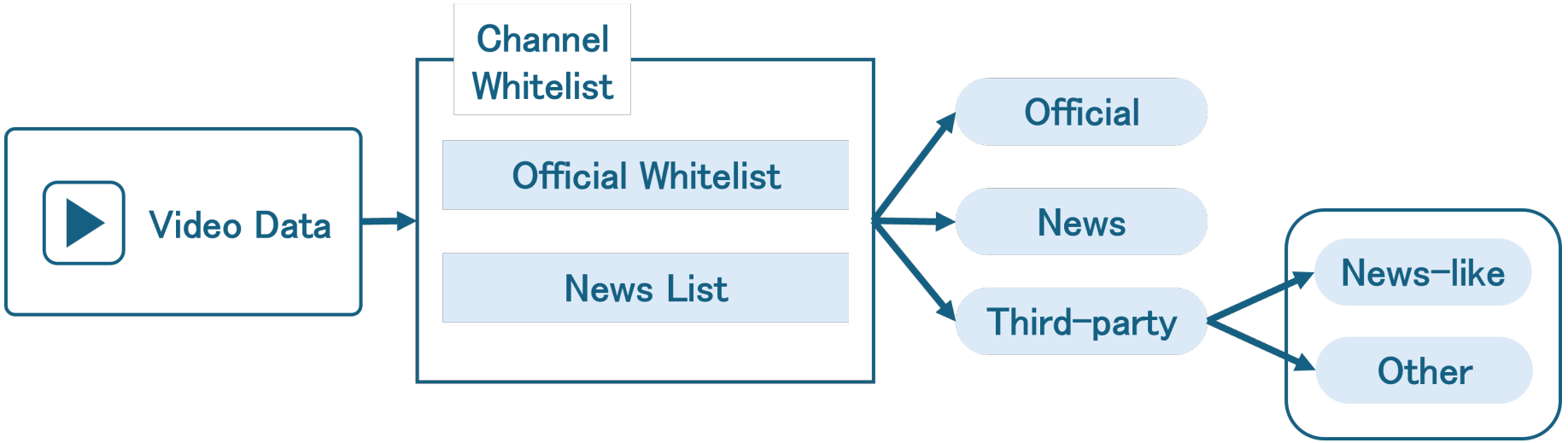
- Election Date: February 14th, 2024
- Official Result Announcement: March 20th, 2024
- Candidates
  ① Anies Baswedan
  ② Prabowo Subianto
  ③ Ganjar Pranowo

Video Data

- Target : YouTube videos under five minutes
- Search Keywords
  - The name of three presidential candidates
  - Indonesian terms for “presidential election” or “election”
- Period
  - From late November 2023 to early June 2024
- Preprocessing
  - Transcribed to text by using an off-the-shell ASR model, Wisper
- Number of videos
  - 36,365 posts

[3] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I.: Robust speech recognition via Large-Scale Weak Supervision. arXiv preprint arXiv:2212.04356 (2022)

Video Classification



- First stage: Classify video posts into three information sources
  - Either “official”, “news”, or “third-party”
  - By using hand-crafted channel names
- Second stage: Classify “third-party” into “News-like” or “Other”
  - By using IndoBERT fine-tuned with pseudo training data

Evaluation of Second Stage Classification

- Classify transcripts of Third-party videos into either “news-line” or “other”.
- Employed pre-trained model IndoBERT.
- Fine-tuned by 9,000 transcripts labeled “News” and “Third-Party” in the first stage classification.
  - regarding “News” as “news-like” and “Thrid-party” as “other”

	test (1,000)
Accuracy	0.9091
Precision	0.4000
Recall	0.2222
F1 Score	0.2857

- The performance was not satisfactory in terms of precision and recall.

→ We decided to focus on “Third-Party” as a whole for further analysis.

Sentiment Analysis

- Three categories
  - **Positive**
  - **Negative**
  - **Neutral**

- Classifier
  - **Indonesian RoBERTa Base Sentiment Classifier [5]**
    → Indonesian RoBERTa based pre-trained model fine-tuned by using Indonesian dataset for sentiment analysis

[5] Wongso, W.: indonesian-roberta-base-sentiment-classifier (Revision e402e46) (2023)

Sentiment Analysis Evaluation

- Investigate how the off-the-shell classifier works well on the YouTube posts.
  - 100 third-party videos were randomly selected, then manually labeled.
  - After excluding 12 videos labeled “undetermined”, 88 were used for evaluation.
- Evaluation Metric: F1 measure
- Result: Accuracy 0.76, F1-score 0.76

	precision	recall	f1-score
positive	**0.97**	0.61	0.75
negative	0.66	**0.94**	**0.78**
neutral	0.67	0.89	0.76

Sentiment Impact Score (SIS) [2]

- A quantification index that evaluate the sentiment and frequency of posts

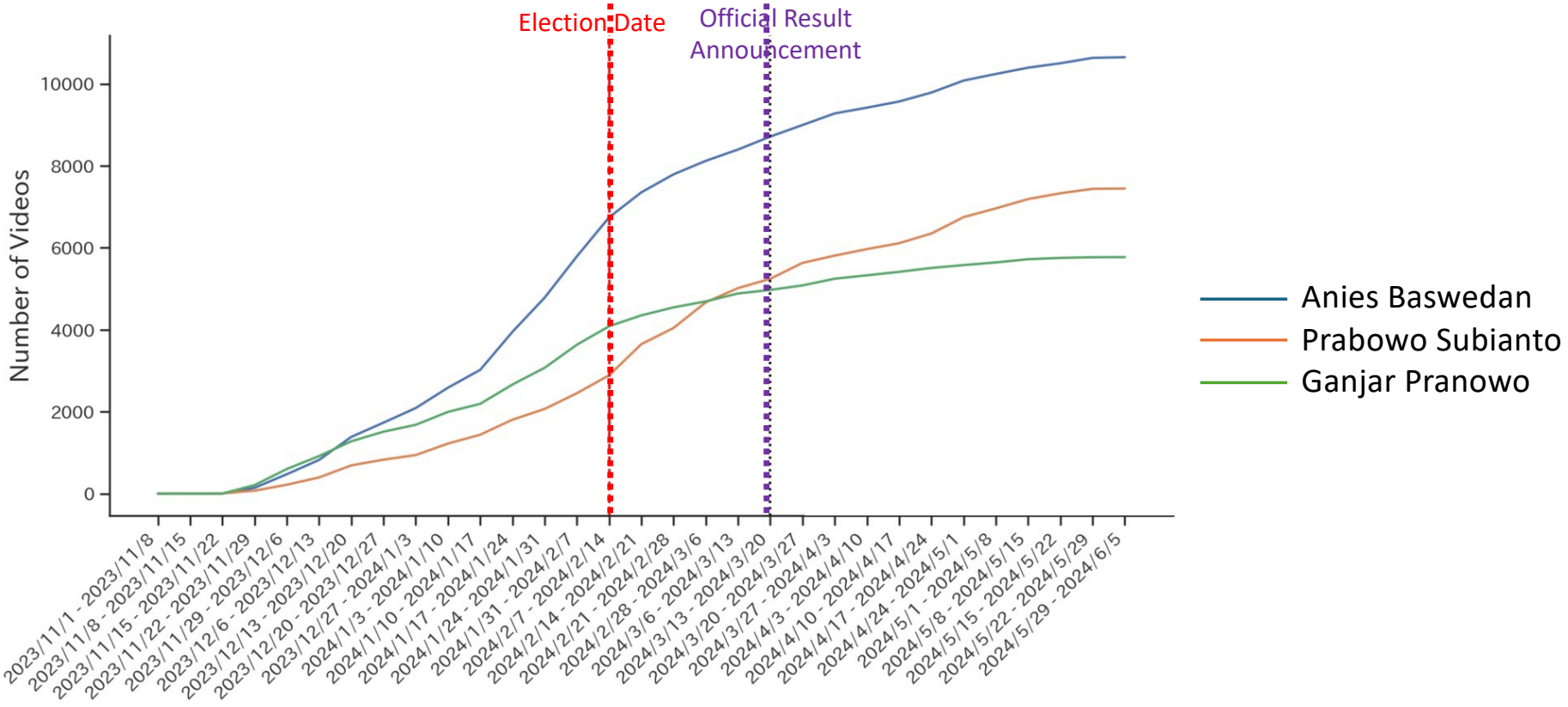
$$SIS = \left(\frac{\omega - \psi}{\phi}\right) \times \log(\phi)$$

  - ω : number of positive posts
  - ψ : number of negative posts
  - ϕ : number of all posts excluding neutral

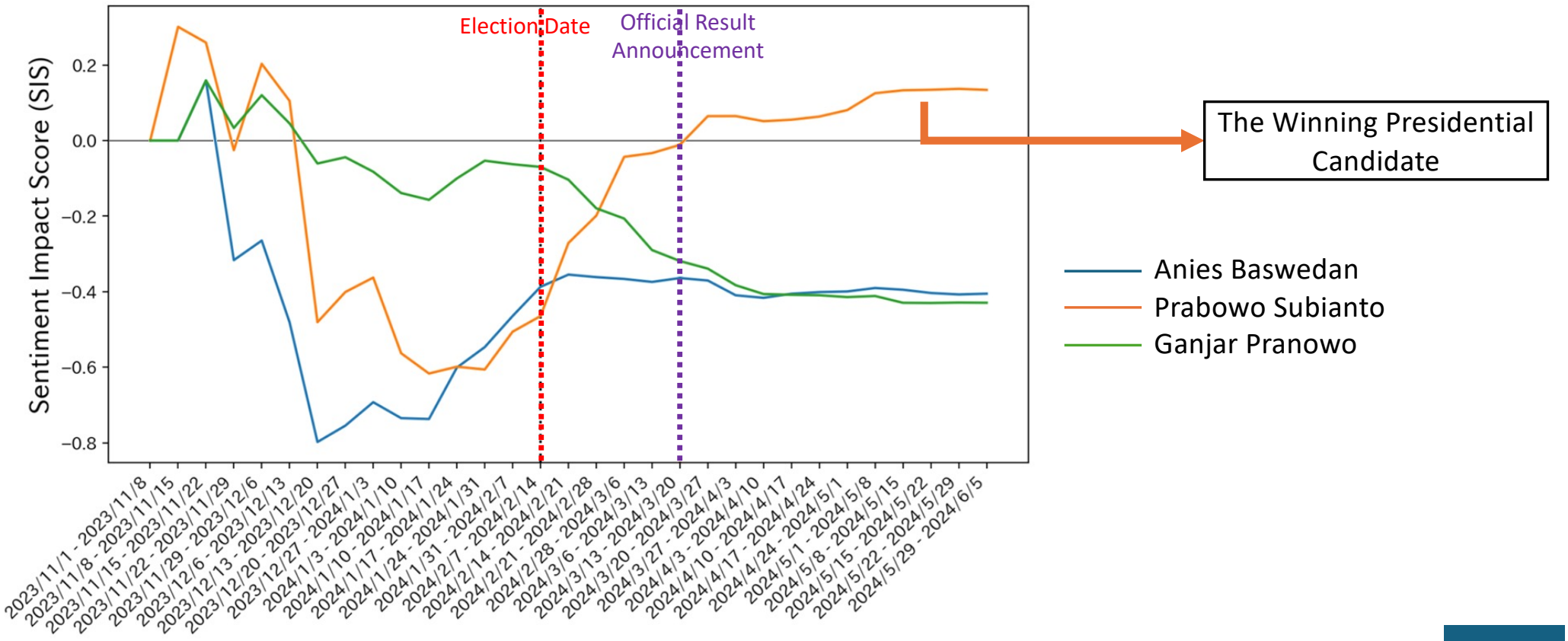
- When $\omega > \psi$, the score becomes a plus number.
- When $\omega < \psi$, the score becomes a minus number.
- Greater media coverage impacts absolute value of the score.

[2] Bovafiz, M.: Leveraging AI and Sentiment Analysis for Forecasting Election Outcomes in Mauritius, arXiv preprint arXiv:2410.20859 (2024)

The Cumulative Number of Videos of Published Third-Party Videos



SIS based on the Cumulative Number of Published Third-Party Video



Conclusion

- We conducted sentiment analysis on YouTube videos and compute Sentiment Impact Score (SIS) to quantify the overall sentiment trends.
- The result shows:
  - Majority of the videos were classified as negative.
  - The sentiment trend of third-party videos shows the shift in public sentiment favoring the elected presidential candidate around the time of election date.