

Public Sentiment Analysis of President and Vice President Candidates in Indonesia in 2024 Via Social Media Platform X

Herman Heriadi¹, Asrul Syam², Risnayanti A Djamro³, Marsa⁴

Informatics Engineering, Universitas Dipa Makasaar, Indonesia

Jalan Perintis Kemerdekaan Km. 9, Tamalanrea Indah, Kota Makassar

e-mail: herman.heriadi@undipa.ac.id, asrulsyam@undipa.ac.id, risnayanti@undipa.ac.id,

marsa@undipa.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap calon presiden dan calon wakil presiden di Indonesia pada tahun 2024. Melalui pengumpulan dan analisis data dari platform media sosial X, penelitian ini merinci pandangan dan perasaan masyarakat terhadap kandidat potensial. Metode analisis sentimen menggunakan naïve bayes untuk mengidentifikasi dan mengevaluasi opini positif, negatif, dan netral yang terkait dengan setiap calon. Hasil penelitian ini diharapkan dapat memberikan wawasan mendalam tentang preferensi publik dan potensi isu-isu yang mungkin memengaruhi pemilihan presiden pada tahun 2024 di Indonesia.

Kata kunci: Machine Learning, Analisis Sentimen, Calon Presiden, Calon Wakil Presiden, Media Sosial X.

Abstract

This research aims to analyze public sentiment towards presidential and vice-presidential candidates in Indonesia in the year 2024. Through the collection and analysis of data from social media platforms X, this study details the views and feelings of the public towards potential candidates. Sentiment analysis methods employ naïve bayes techniques to identify and evaluate positive, negative, and neutral opinions associated with each candidate. The findings of this research are expected to provide profound insights into public preferences and potential issues that may influence the presidential election in 2024 in Indonesia.

Keywords: Machine Learning, Presidential Candidates, Sentiment Analysis, Vice Presidential Candidates, Social Media X.

1. Pendahuluan

Indonesia's presidential and vice presidential elections are one of the crucial moments in the country's democratic life [1]. In 2024, Indonesia will again face a democratic challenge in determining the leader who will lead the country for the next period. One important aspect of this election process is understanding public sentiment toward presidential and vice presidential candidates.

People often use social media to express their opinions, feelings, experiences, or other things that concern them [2]. Therefore, this study aims to conduct an in-depth analysis of public sentiment regarding Indonesia's presidential and vice presidential candidates in 2024. By utilizing the social media platform X, this study is expected to provide an overview of public perception of potential candidates. Tweets on the X platform are believed to be accurate data because public opinion is formed from an outburst of emotion, opinions that arise, or stories that are going viral on social media [3].

The sentiment analysis method using the Naïve Bayes technique will be used to identify and evaluate positive, negative, and neutral opinions that emerge from various sources. Sentiment Analysis is the process of understanding and processing textual data automatically to obtain sentiment information contained in a sentence or text in the form of an opinion [4].

The naïve Bayes Classifier is a classification method that is rooted in Bayes' Theorem [5]. The classification method uses probability and statistical methods, which are predicting opportunities based on previous experience (Bayes' Theorem) with its main characteristic being a very strong (naive) assumption of the dependence of each condition/event [6].

Thus, a better understanding of public sentiment can serve as a basis for more informed and data-driven decision-making in formulating campaign strategies and policies that support the country's progress and stability.

2. Metode Penelitian

2.1. Data Collection

There are three approaches described for collecting data from social media platforms. First, this method requires downloading data directly from a database server, this requires close collaboration with the social media company. Second, an alternative method is to collect data through a software programming interface, also known as an API (Application Programming Interface). An API is a set of commands that allow software developers to interact with other software [7]. Software developers can use pre-coded scripts, which usually require permission from the social media platform owner to use them. The third method uses web scraping, which is the process of extracting valuable and interesting text information from web pages [8]. This method is especially useful for social media platforms that do not have an API. This is done using the HTTP text transfer protocol. When it comes to data collection, elements such as method, time, population, and sample are crucial to the overall process.

The following are the steps for data collection in this study:

- Installing WebHarvy Evaluation for web scraping.
- Scraping twitter data on posts on social media platform X.
- Exporting comments to a .csv file

Based on the time of data collection in this study, it is included in the Cross Section approach, the data will be collected over a certain period [9]. In this study, time is needed to collect data throughout the First Presidential Candidate Debate (December 12, 2023), the Second Vice Presidential Candidate Debate (December 22, 2023), the Third Presidential Candidate Debate (January 7, 2024), and the Fourth Vice Presidential Candidate Debate (January 21, 2024).

Data collection for Presidential Candidate Anies Rasyid Baswedan through the X social media platform using keywords, Anies, Anies Baswedan, Presidential Candidate 01, Presidential Candidate 1, Candidate Pair 01, and Candidate Pair 1. By using the keywords above, the researcher obtained a tweet dataset that the researcher displays in the table below.

Table 1. Dataset Presidential Candidate Anies Rasyid Baswedan

Dataset Anies Rasyid Baswedan
Harun Al-Rasyid, KM50 dan Kanjuruhan disinggung oleh Capres Paslon No 1, Anies Baswedan: Ini harus dituntaskan, ini harus bisa menghadirkan ras keadilanâ€ Cuma Pak Anies yg angkat kasus itu, tapi herannya pendukung PS bersikap sinis dan tanya knp bawaÂ² orang mati, kan aneh. Setelah debat capres semalem update odd menjadi Anis 2.34 hdp 1-1.5 Prabowo 3.01 hdp 0.5-1 Ganjar 5.12 0.5 BTTS Anis lolos putaran 2 Prabowo lolos putaran 2 Ganjar tidak lolos 2Ã—45 Anis/anis Seri/anis Seri/prabowo Prabowo/kalah ganjar/seri Seri/kalah. Bismillah. Dari hasil debat capres pertama mantap dgn paslon no 1. Ditambah dengan sudah ditandatangani pakta integritas, makin yakin dengan no 1. Kita buktikan bahwa pak anies orang yg amanah. Kalian gimana?

Data collection for Presidential Candidate Prabowo Subianto through social media platform X using keywords, Prabowo, Subianto, Presidential Candidate 02, Presidential Candidate 2, Candidate Pair 02 and Candidate Pair 2. By using the keywords above, the researcher obtained a tweet dataset that the researcher displays in the table below.

Table 2. Dataset Presidential Candidate Prabowo Subianto

Dataset Prabowo Subianto
Prabowo hampir bisa dipastikan sulit dikalahkan. Bahkan target Paslon 02 sudah di level menang 1 putaran. Makanya sangat wajar ketika 01 dan 03 lebih memilih untuk menyerang 02. 01 dan 03 sulit Juara 1 di Putaran 1. Namun keduanya mungkin bisa menang jika bisa masuk Putaran 2.â€ Reaksi pendukung paslon 02 usia nobar debat capres. Mari bersama sama Kita AAMMIIN-KAN. Semoga Allah SWT, Tuhan Pemilik Kuasa Kabulkan Permintaan Prabowo agar Rakyat Indonesia TIDAK PILIH MEREKA PASLON 02. Aaammiiin Ya Robb.

Data collection for Presidential Candidate Ganjar Pranowo through social media platform X using keywords, Ganjar, Pranowo, Presidential Candidate 03, Presidential Candidate 3, Candidate Pair 03, and Candidate Pair 3. By using the keywords above, the researcher obtained a tweet dataset that the researcher displays in the table below.

Table 3. Dataset Presidential Candidate Ganjar Pranowo

Dataset Ganjar Pranowo
Saya yakin setelah nonton #debatcapres, besok akan rilis survey kalau capres No 2 akan meraih 97% kemenangan telak Paslon 1 = 2% Capres No 3 = 1% Yang di survey adalah: 1. Capres yg emosian 2. Capres yg joget kl gak bs jawab 3. Capres yg endors Jokowi mulu. #DebatCapres2023 Paslon 1 Anies Rasyid Baswedan tidak kalah penting Hadir orang tua almarhum ananda Harun Al Rasyid Di tengah capres No 2 dan No 3 Korban meninggal dunia tragedi 22 mei 2019 Bapak Didin Wahyudin masih berharap keadilan, usut tuntas #DebatCapres Next KM 50. Survei Poltracking Terbaru Elektabilitas 3 Paslon Pipres 2024 Prabowo-Gibran 45,2% Ganjar-Mahfud 27,3% Anies-Imin 23,1% TT/TJ 4,4% Paslon 01 & 03 silahkan kalian cakar2an berebut 1 tiket putaran kedua...klo paslon 02 kerja lebih keras agar menang 1 putaran.

Data collection on Vice Presidential Candidate Muhaimin Iskandar via social media X using keywords, Muhaimin Iskandar, Cak Imin, Gus Imin, Gus Muhaimin, Vice Presidential Candidate 01, Vice Presidential Candidate 1. By using the keywords above, the researcher obtained a tweet dataset which the researcher displays in the table below.

Table 4. Dataset Vice Presidential Candidate Muhaimin Iskandar

Dataset Muhaimin Iskandar
Mas Gibran : "dulu pak muhaimin pernah ikut potong tumpeng di IKN, kenapa jadi ngga konsisten gini" Yahhh kenapa tuh slepet #DebatCawapres Cak Imin. El Sulfat : " gus muhaimin agak aneh ya pengen bangun kota selevel jakarta tapi tidak setuju sama ikn , Tapi ya monggo lah terserah" Biarkann EL SULFAT MEMASAK #DebatCawapres Cak Imin. Paslon Nomor Urut 1, Anies Baswedan dan Muhaimin Iskandar (Cak Imin) menyampaikan keterangannya usai debat Cawapres pada Jumat, (22/12/2023). Anies dan Cak Imin sempat mengomentari strategi dari pihak lain dalam memberikan pertanyaan saat debat. Diketahui Debat Cawapres untukâ€¦

Data collection for Vice Presidential Candidate Gibran Rakabuming through social media platform X using keywords, Gibran Rakabuming, Samsul, Sulfat, Vice Presidential Candidate 02, and Vice Presidential Candidate 2. By using the keywords above, the researcher obtained a tweet dataset which the researcher displays in the table below.

Table 5. Dataset Vice Presidential Candidate Gibran Rakabuming

Dataset Gibran Rakabuming
Sesi Satu Debat Cawapres 1. AMIN : Cak Imin, dengan slepetnya. 7/10 2. Prabowo Gibran : Gibran dengan gagasan hilirisasi digital, 7,5/10 3. GAMA : Pak Mahfud dengan proyeksi pertumbuhan 7% di masa depan, 7,5/10. Usai 2 debat pertama antara para capres-cawapres. Siapa kira-kira pilihan anda? Secara Overall, Nilai untuk Debat Cawapres malam ini 1. Cak Imin : 73 2. Gibran : 69 3. Mahfud MD : 72. Setuju engga? PS: Gibran nilainya dikurangi karena ada prilaku songong sama orang tua.

Data collection for Vice Presidential Candidate Mahfud MD through social media platform X using keywords, Mahfud MD, Prof Mahfud, Vice Presidential Candidate 03, and Vice Presidential Candidate 3. By using the keywords above, the researcher obtained a tweet dataset which the researcher displays in the table below.

Table 6. Dataset Vice Presidential Candidate Mahfud MD

Dataset Mahfud MD
Bicara soal hukum dan ekonomi..??Cocok Banget dg Kapasitas Cawapres No 3 Prof @mohmahfudmd Seperti yg dijelaskan oleh banyak Pihak bahwa ekonomi yg kuat dibangun diatas pondasi hukum yang kuat, dan sosok itu adalah Prof Mahfud MD si Pendekar Hukum 3konomi #Hukum3konomiMahfud. Sambil Nonton DebatCaawapres yukk simak apa yang di harapkan oleh Cawapres No 3 Prof @mohmahfudmd Beliau berharap ada nya SSK yg di bentuk oleh LPSK kepastian hukum nya harus benar benar di Tegakan...!! #SahabatGanjar Ganjar Mahfud Satu Set. Cawapres No 3 Prof Mahfud MD: Jangan Meminta Atau Memburu Jabatan,Karna Itu Sama Saja Kita Bekerja Buat Diri Sendiri,Bkn Untuk Rakyat.

2.1. Pre-Processing

After successfully collecting data from social media platforms, the next crucial step is the pre-processing stage. Data pre-processing plays an important role in ensuring the accuracy, cleanliness, and regularity of the data that has been collected. Here are the stages in data pre-processing:

a. Cleaning

Cleaning is the process of removing certain elements, such as duplicate data, emojis, special characters, numbers, three-letter words, and excess spaces, by applying certain strategies [10]. Cleaning is

the first step to ensure the cleanliness and accuracy of the dataset. This process involves removing anomalies, duplications, and missing values to make the data more consistent and reliable.

Then, case folding is a stage that equates all texts that will be entered into the model to lowercase [11]. Through case folding, all characters in the text are converted to lowercase, thus eliminating ambiguities that may arise due to capitalization variations. This process facilitates consistent analysis and reduces complexity in text processing.

b. Normalization

Data normalization is a method used to overcome problems and constraints related to data ranges that are too far apart [12]. In addition, normalization is the process of transforming data to conform to certain standards. This process involves forming data in a uniform format, allowing for easier comparison and analysis. For example, the word "yg" will be changed to "yang" and "gk" will be changed to "tidak".

c. Stopwords

The stopwords deleting stage is a critical process in text processing. Stopwords are common words that often do not provide significant meaning in the analysis [13]. Stopwords deletion can increase focus on more informative and relevant keywords for understanding the content. Examples of words listed in the Indonesian stopwords database are and, by, or this.

d. Tokenization

Tokenization is the process of breaking text into smaller units, called tokens [14]. Tokens can be words, phrases, or characters depending on the analysis needs. Tokenization supports further processing such as search, classification, and sentiment analysis.

e. Stemming

Stemming is needed to reduce the variation of indexes in data so that words that have a prefix or suffix will be simplified to their basic form [15]. In stemming, the basic form of a word is identified by removing the suffix or prefix of the word. The purpose of stemming is to reduce words to their basic form so that morphological variations do not interfere with the analysis. For example, the word "disinggung" can be simplified to "singgung".

f. Data Analysis

Data analysis is a critical step in gaining meaningful insights. By involving statistical techniques, machine learning methods such as Naïve Bayes, or other approaches, data analysis provides a deep understanding of patterns, trends, and relationships in the dataset. The results of this analysis can be used to make informed decisions or make predictions such as negative, positive, or neutral sentiments.

3. Hasil dan Pembahasan

This study uses the Naïve Bayes method to analyze public sentiment toward presidential and vice presidential candidates in Indonesia in 2024. After going through the sentiment analysis process, the following are the results obtained for the presidential and vice presidential candidates.

a. Cleaning

For the cleaning process, please see the table below.

Table 7. Celaning Dataset Presidential Candidate Anis Baswedan

Cleaning
harun alrasyid km50 dan kanjuruhan disinggung oleh capres paslon no 1 anies baswedan ini harus dituntaskan ini harus bisa menghadirkan ras keadilan cuma pak anies yg angkat kasus itu tapi herannya pendukung ps bersikap sinis dan tanya knp bawa orang mati kan aneh.
setelah debat capres semalem update odd menjadi anis 234 hdp 115 prabowo 301 hdp 051 ganjar 512 05 btts anis lolos putaran 2 prabowo lolos putaran 2 ganjar tidak lolos 245 anisanis serianis seriprabowo prabowokalah ganjarseri serikalah.
bismillah dari hasil debat capres pertama mantap dgn paslon no 1 ditambah dengan sudah ditandatanganinya pakta integritas makin yakin dengan no 1 kita buktikan bahwa pak anies orang yg amanah kalian gimana.

Table 8. Celaning Dataset Presidential Candidate Prabowo Subianto

Cleaning
prabowo hampir bisa dipastikan sulit dikalahkan bahkan target paslon 02 sudah di level menang 1 putaran makanya sangat wajar ketika 01 dan 03 lebih memilih untuk menyerang 02 01 dan 03 sulit juara 1 di putaran 1 namun keduanya mungkin bisa menang jika bisa masuk putaran 2.
reaksi pendukung paslon 02 usia nobar debat capres.
mari bersama sama kita aammiinkan semoga allah swt tuhan pemilik kuasa kabulkan permintaan prabowo agar rakyat indonesia tidak pilih mereka paslon 02 aaammiin ya robb.

Table 9. Celaning Dataset Presidential Candidate Ganjar Pranowo

Cleaning
saya yakin setelah nonton besok akan rilis survey kalau capres no 2 akan meraih 97 kemenangan telak paslon 1 2 capres no 3 1 yang di survey adalah 1 capres yg emosian 2 capres yg joget kl gak bs jawab 3 capres yg endors jokowi mulu. paslon 1 anies rasyid baswedan tidak kalah penting hadir orang tua almarhum ananda harun al rasyid di tengah capres no 2 dan no 3 korban meninggal dunia tragedi 22 mei 2019 bapak didin wahyudin masih berharap keadilan usut tuntas next km 50. survei poltracking terbaru elektabilitas 3 paslon pipres 2024 prabowogibran 452 ganjarmahfud 273 aniesimin 231 ttj 44 paslon 01 03 silahkan kalian cakar2an berebut 1 tiket putaran kedua klo paslon 02 kerja lebih keras agar menang 1 putaran.

Table 10. Celaning Dataset VicePresidential Candidate Muhaimin Iskandar

Cleaning
mas gibran dulu pak muhaimin pernah ikut potong tumpeng di ikn kenapa jadi ngga konsisten gini yahhh kena tuh slepet cak imin. el sulfat gus muhaimin agak aneh ya pengen bangun kota selevel jakarta tapi tidak setuju sama ikn tapi ya monggo lah terserah biarkan el sulfat memasak cak imin. paslon nomor urut 1 anies baswedan dan muhaimin iskandar cak imin menyampaikan keterangannya usai debat cawapres pada jumat 22/2/2023 anies dan cak imin sempat mengomentari strategi dari pihak lain dalam memberikan pertanyaan saat debat diketahui debat cawapres untuk.

Table 11. Celaning Dataset Vice Presidential Candidate Gibran Rakabuming

Cleaning
sesi satu debat cawapres 1 amin cak imin dengan slepetnya 710 2 prabowo gibran gibran dengan gagasan hilirisasi digital 7510 3 gama pak mahfud dengan proyeksi pertumbuhan 7 di masa depan 7510. usai 2 debat pertama antara para caprescawapres siapa kirakira pilihan anda. secara overall nilai untuk debat cawapres malam ini 1 cak imin 73 2 gibran 69 3 mahfud md 72 setuju engga ps gibran nilainya dikurangi karena ada prilaku songong sama orang tua.

Table 12. Celaning Dataset Vice Presidential Candidate Mahfud MD

Cleaning
bicara soal hukum dan economicocok banget dg kapasitas cawapres no 3 prof seperti yg dijelaskan oleh banyak pihak bahwa ekonomi yg kuat dibangun diatas pondasi hukum yg kuat dan sosok itu adalah prof mahfud md si pendekar hukum 3konomi. sambil nonton debatcaawapres yukk simak apa yang di harapkan oleh cawapres no 3 prof beliau berharap ada nya ssk yg di bentuk oleh lpsk kepastian hukum nya harus benar benar di tegakan ganjar mahfud satu set. Cawapres no 3 prof mahfud md jangan meminta atau memburu jabatankarna itu sama saja kita bekerja buat diri sendiriibkn untuk rakyat.

b. Normalization

The normalization process can be seen in the table below.

Table 13. Normalization Dataset Presidential Candidate Anis Baswedan

Normalization
harun alrasyid km50 dan kanjuruhan disinggung oleh capres paslon nomor 1 anies baswedan ini harus dituntaskan ini harus bisa menghadirkan ras keadilan cuma pak anies yang angkat kasus itu tetapi herannya pendukung ps bersikap sinis dan tanya kenapa bawa orang mati kan aneh setelah debat capres semalem update odd menjadi anis 234 hdp 115 prabowo 301 hdp 051 ganjar 512 05 btts anis lolos putaran 2 prabowo lolos putaran 2 ganjar tidak lolos 245 anisanis serianis seriprabowo prabowokalah ganjarseri serikalah bismillah dari hasil debat capres pertama mantap dgn paslon nomor 1 ditambah dengan sudah ditandatangani pakta integritas makin yakin dengan nomor 1 kita buktikan bahwa pak anies orang yang amanah kalian bagaimana

Table 14. Normalization Dataset Presidential Candidate Prabowo Subianto

Normalization
prabowo hampir bisa dipastikan sulit dikalahkan bahkan target paslon 02 sudah di level menang 1 putaran makanya sangat wajar ketika 01 dan 03 lebih memilih untuk menyerang 02 01 dan 03 sulit juara 1 di putaran 1 namun keduanya mungkin bisa menang jika bisa masuk putaran 2. reaksi pendukung paslon 02 usia nobar debat capres. mari bersama sama kita aammiinkan semoga allah swt tuhan pemilik kuasa kabulkan permintaan prabowo agar rakyat indonesia tidak pilih mereka paslon 02 aaammiin ya robb.

Table 15. Normalization Dataset Presidential Candidate Ganjar Pranowo

Normalization
saya yakin setelah nonton besok akan rilis survey kalau capres nomor 2 akan meraih 97 kemenangan telak paslon 1 2 capres nomor 3 1 yang di survey adalah 1 capres yang emosian 2 capres yang joget kl tidak bisa jawab 3 capres yang

endors jokowi mulu.
 paslon 1 anies rasyid baswedan tidak kalah penting hadir orang tua almarhum ananda harun al rasyid di tengah capres nomor 2 dan nomor 3 korban meninggal dunia tragedi 22 mei 2019 bapak didin wahyudin masih berharap keadilan usut tuntas next km 50.
 survei poltracking terbaru elektabilitas 3 paslon pipres 2024 prabowogibran 452 ganjarmahfud 273 aniesimin 231 ttj 44 paslon 01 03 **silakan** kalian cakar2an berebut 1 tiket putaran kedua klo paslon 02 kerja lebih keras agar menang 1 putaran.

Table 16. Normalization Dataset Vice Presidential Candidate Muhaimin Iskandar

Normalization

mas gibran dulu pak muhaimin pernah ikut potong tumpeng di ikn kenapa jadi **tidak** konsisten gini yahhh kena tuh slepet cak imin.
 el sulfat gus muhaimin agak aneh ya pengen bangun kota selevel jakarta **tetapi** tidak setuju sama ikn **tetapi** ya monggo lah terserah biarkan el sulfat memasak cak imin.
 paslon nomor urut 1 anies baswedan dan muhaimin iskandar cak imin menyampaikan keterangannya usai debat cawapres pada jumat 22122023 anies dan cak imin sempat mengomentari strategi dari pihak lain dalam memberikan pertanyaan saat debat diketahui debat cawapres untuk.

Table 17. Normalization Dataset Vice Presidential Candidate Gibran Rakabuming

Normalization

sesi satu debat cawapres 1 amin cak imin dengan slepetnya 710 2 prabowo gibran gibran dengan gagasan hilirisasi digital 7510 3 gama pak mahfud dengan proyeksi pertumbuhan 7 di masa depan 7510.
 usai 2 debat pertama antara para capres cawapres siapa kirakira pilihan anda.
 secara overall nilai untuk debat cawapres malam ini 1 cak imin 73 2 gibran 69 3 mahfud md 72 setuju engga ps gibran nilainya dikurangi karena ada perilaku songong sama orang tua.

Table 18. Normalization Dataset Vice Presidential Candidate Mahfud MD

Normalization

bicara soal hukum dan economicocok banget **dengan** kapasitas cawapres **nomor** 3 prof seperti **yang** dijelaskan oleh banyak pihak bahwa ekonomi **yang** kuat dibangun diatas pondasi hukum **yang** kuat dan sosok itu adalah prof mahfud md si pendekar hukum 3konomi.
 sambil nonton debataawapres yuk simak apa yang di harapkan oleh cawapres **nomor** 3 prof beliau berharap ada nya ssk yang di bentuk oleh lpsk kepastian hukum nya harus benar benar di tegakan ganjar mahfud satu set.
 cawapres **nomor** 3 prof mahfud md jangan meminta atau memburu jabatankarna itu sama saja kita bekerja buat diri sendiri bkn untuk rakyat.

c. Stopwords

The stopword process can be seen in the following table.

Table 19. Stopwords Dataset Presidential Candidate Anis Baswedan

Stopwords

harun alrasyid km50 kanjuruhan disinggung capres paslon 1 anies baswedan harus dituntaskan harus menghadirkan ras keadilan Cuma pak anies angkat kasus tetapi herannya pendukung ps bersikap sinis tanya bawa orang mati kan aneh debat capres semalem update odd menjadi anis 234 hdp 115 prabowo 301 hdp 051 ganjar 512 05 bts anis lolos putaran 2 prabowo lolos putaran 2 ganjar lolos 245 anisanis serianis seriprabowo prabowokalah ganjarseri serikalah bismillah hasil debat capres pertama mantap dgn paslon 1 ditambah sudah ditandatangani pakta integritas makin yakin 1 buktikan pak anies orang amanah kalian bagaimana

Table 20. Stopwords Dataset Presidential Candidate Prabowo Subianto

Stopwords

prabowo hampir dipastikan sulit dikalahkan bahkan target paslon 02 level menang 1 putaran makanya sangat wajar 01 03 lebih memilih menyerang 02 01 03 sulit juara 1 di putaran 1 keduanya mungkin menang bisa masuk putaran 2 reaksi pendukung paslon 02 usia nobar debat capres bersama sama aammiinkan semoga allah swt tuhan pemilik kuasa kabulkan permintaan prabowo rakyat indonesia pilih paslon 02 aammiin robb

Table 21. Stopwords Dataset Presidential Candidate Ganjar Pranowo

Stopwords

yakin nonton besok rilis survey kalau capres 2 meraih 97 kemenangan telak paslon 1 2 capres 3 1 di survey 1 capres emosian 2 capres joget kl bisa jawab 3 capres endors jokowi mulu
 paslon 1 anies rasyid baswedan kalah penting hadir orang tua almarhum ananda harun al rasyid tengah capres 2 3 korban meninggal dunia tragedi 22 mei 2019 bapak didin wahyudin berharap keadilan usut tuntas next km 50
 survei poltracking terbaru elektabilitas 3 paslon pipres 2024 prabowogibran 452 ganjarmahfud 273 aniesimin 231 ttj 44 paslon 01 03 silakan kalian cakar2an berebut 1 tiket putaran kedua klo paslon 02 kerja lebih keras menang 1 putaran

Table 22. Stopwords Dataset Vice Presidential Candidate Muhaimin Iskandar

Stopwords
mas gibran dulu pak muhaimin pernah ikut potong tumpeng ikn jadi konsisten gini yahhh kena tuh slepet cak i min el sulfat gus muhaimin aneh pengen bangun kota selevel jakarta tidak setuju sama ikn ya monggo lah terserah biarkann el sulfat memasak cak imin paslon nomor urut 1 anies baswedan muhaimin iskandar cak imin menyampaikan keterangannya usai debat cawapres jumat 22122023 anies cak imin sempat mengomentari strategi pihak dalam memberikan pertanyaan debat diketahui debat cawapres

Table 23. Stopwords Dataset Vice Presidential Candidate Gibran Rakabuming

Stopwords
sesi satu debat cawapres 1 amin cak imin slepetnya 710 2 prabowo gibran gibran gagasan hilirisasi digital 7510 3 gama pak mahfud proyeksi pertumbuhan 7 masa depan 7510 usai 2 debat pertama para caprescawapres siapa kirakira pilihan overall nilai debat cawapres malam 1 cak imin 73 2 gibran 69 3 mahfud md 72 setuju engga ps gibran nilainya dikurangi ada prilaku songong sama orang tua

Table 24. Stopwords Dataset Vice Presidential Candidate Mahfud MD

Stopwords
bicara soal hukum economicocok banget kapasitas cawapres 3 prof dijelaskan banyak pihak ekonomi kuat dibangun diatas pondasi hukum yang kuat sosok adalah prof mahfud md si pendekar hukum 3konomi nonton debatacawapres yukk simak apa harapkan cawapres 3 prof beliau berharap nya ssk di bentuk lpsr kepastian hukum nya benar benar di tegakan ganjar mahfud satu set cawapres 3 prof mahfud md jangan meminta memburu jabatankarna sama kita bekerja buat diri sendiribkn rakyat

d. Tokenization

The tokenization process can be seen in the table below.

Table 25. Stopwords Dataset Presidential Candidate Anis Baswedan

Tokenization
[harun', 'alrasyid', 'km50', 'kanjuruhan', 'disinggung', 'capres', 'paslon', '1', 'anies', 'baswedan', 'harus', 'dituntaskan', 'harus', 'menghadirkan', 'ras', 'keadilan', 'cuma', 'pak', 'anies', 'angkat', 'kasus', 'tetapi', 'herannya', 'pendukung', 'ps', 'bersikap', 'sinis', 'tanya', 'bawa', 'orang', 'mati', 'kan', 'aneh'] [debat', 'capres', 'semalem', 'update', 'odd', 'menjadi', 'anis', '234', 'hdp', '115', 'prabowo', '301', 'hdp', '051', 'ganjar', '512', '05', 'btts', 'anis', 'lolos', 'putaran', '2', 'prabowo', 'lolos', 'putaran', '2', 'ganjar', 'lolos', '245', 'anisanis', 'serianis', 'seriprabowo', 'prabowokalah', 'ganjarseri', 'serikalalah'] [bismillah', 'hasil', 'debat', 'capres', 'pertama', 'mantap', 'dgn', 'paslon', '1', 'ditambah', 'sudah', 'ditandatangani', 'pakta', 'integritas', 'makin', 'yakin', 'tidak', '1', 'buktikan', 'pak', 'anies', 'orang', 'amanah', 'kalian', 'bagaimana']

Table 26. Stopwords Dataset Presidential Candidate Prabowo Subianto

Tokenization
['prabowo', 'hampir', 'dipastikan', 'sulit', 'dikalahkan', 'bahkan', 'target', 'paslon', '02', 'level', 'menang', '1', 'putaran', 'makanya', 'sangat', 'wajar', '01', '03', 'lebih', 'memilih', 'menyerang', '02', '01', '03', 'sulit', 'juara', '1', 'di', 'putaran', '1', 'keduanya', 'mungkin', 'menang', 'bisa', 'masuk', 'putaran', '2'] [reaksi', 'pendukung', 'paslon', '02', 'usia', 'nobar', 'debat', 'capres'] [bersama', 'sama', 'aammiinkan', 'semoga', 'allah', 'swt', 'tuhan', 'pemilik', 'kuasa', 'kabulkan', 'permintaan', 'prabowo', 'rakyat', 'indonesia', 'pilih', 'paslon', '02', 'aaammiin', 'robb']

Table 27. Stopwords Dataset Presidential Candidate Ganjar Pranowo

Tokenization
[yakin', 'nonton', 'besok', 'rilis', 'survey', 'kalau', 'capres', '2', 'meraih', '97', 'kemenangan', 'telak', 'paslon', '1', '2', 'capres', '3', '1', 'di', 'survey', '1', 'capres', 'emosian', '2', 'capres', 'joget', 'kl', 'bisa', 'jawab', '3', 'capres', 'endors', 'jokowi', 'mulu'] [paslon', '1', 'anies', 'rasyid', 'baswedan', 'kalah', 'penting', 'hadir', 'orang', 'tua', 'almarhum', 'ananda', 'harun', 'al', 'rasyid', 'tengah', 'capres', '2', 'tidak', '3', 'korban', 'meninggal', 'dunia', 'tragedi', '22', 'mei', '2019', 'bapak', 'didin', 'wahyudin', 'berharap', 'keadilan', 'usut', 'tuntas', 'next', 'km', '50'] [survei', 'poltracking', 'terbaru', 'elektabilitas', '3', 'paslon', 'pipres', '2024', 'prabowogibran', '452', 'ganjarmahfud', '273', 'aniesimin', '231', 'tttj', '44', 'paslon', '01', '03', 'silakan', 'kalian', 'cakar2an', 'berebut', '1', 'tiket', 'putaran', 'keduaklo', 'paslon', '02', 'kerja', 'lebih', 'keras', 'menang', '1', 'putaran']

Table 28. Stopwords Dataset Vice Presidential Candidate Muhaimin Iskandar

Tokenization
[mas', 'gibran', 'dulu', 'pak', 'muhaimin', 'pernah', 'ikut', 'potong', 'tumpeng', 'ikn', 'jadi', 'konsisten', 'gini', 'yahhh', 'kenaa', 'tuh', 'slepet', 'cak', 'imin'] [el', 'sulfat', 'gus', 'muhaimin', 'aneh', 'pengen', 'bangun', 'kota', 'selevel', 'jakarta', 'tidak', 'setuju', 'sama', 'ikn', 'ya', 'monggo', 'lah', 'terserah', 'biarkann', 'el', 'sulfat', 'memasak', 'cak', 'imin'] [paslon', 'nomor', 'urut', '1', 'anies', 'baswedan', 'muhaimin', 'iskandar', 'cak', 'imin', 'menyampaikan', 'keterangannya',

'usai', 'debat', 'cawapres', 'jumat', '22122023', 'anies', 'cak', 'imin', 'sempat', 'mengomentari', 'strategi', 'pihak', 'dalam', 'memberikan', 'pertanyaan', 'debat', 'diketahui', 'debat', 'cawapres']

Table 29. Stopwords Dataset Vice Presidential Candidate Gibran Rakabuming

Tokenization
['sesi', 'satu', 'debat', 'cawapres', '1', 'amin', 'cak', 'imin', 'slepetnya', '710', '2', 'prabowo', 'gibran', 'gibran', 'gagasan', 'hilirisasi', 'digital', '7510', '3', 'gama', 'pak', 'mahfud', 'proyeksi', 'pertumbuhan', '7', 'masa', 'depan', '7510']
['usai', '2', 'debat', 'pertama', 'para', 'caprescawapres', 'siapa', 'kirakira', 'pilihan']
['overall', 'nilai', 'debat', 'cawapres', 'malam', '1', 'cak', 'imin', '73', '2', 'gibran', '69', '3', 'mahfud', 'md', '72', 'setuju', 'engga', 'ps', 'gibran', 'nilainya', 'dikurangi', 'ada', 'prilaku', 'songong', 'sama', 'orang', 'tua']

Table 30. Stopwords Dataset Vice Presidential Candidate Mahfud MD

Tokenization
['bicara', 'soal', 'hukum', 'ekonomicocok', 'banget', 'kapasitas', 'cawapres', '3', 'prof', 'dijelaskan', 'banyak', 'pihak', 'ekonomi', 'kuat', 'dibangun', 'diatas', 'pondasi', 'hukum', 'yang', 'kuat', 'sosok', 'adalah', 'prof', 'mahfud', 'md', 'si', 'pendekar', 'hukum', '3konomi']
['nonton', 'debatcawapres', 'yukk', 'simak', 'apa', 'harapkan', 'cawapres', '3', 'prof', 'beliau', 'berharap', 'nya', 'ssk', 'di', 'bentuk', 'lpsk', 'kepastian', 'hukum', 'nya', 'benar', 'benar', 'di', 'tegakan', 'ganjar', 'mahfud', 'satu', 'set']
['cawapres', '3', 'prof', 'mahfud', 'md', 'jangan', 'meminta', 'memburu', 'jabatankarna', 'sama', 'kita', 'bekerja', 'buat', 'diri', 'sendiribkn', 'rakyat']

e. Stemming

Table 31. Stopwords Dataset Presidential Candidate Anis Baswedan

Stemming
harun alrasyid km50 kanjuruhan snggung capres paslon 1 anies baswedan harus tuntas harus hadir ras adil cuma pak anies angkat kasus tetapi heran dukung ps sikap sinis tanya bawa orang mati kan aneh debat capres semalem update odd jadi anis 234 hdp 115 prabowo 301 hdp 051 ganjar 512 05 btts anis lolos putar 2 prabowo lolos putar 2 ganjar lolos 245 anisanis rian seriprabowo prabowokalah ganjarseri serikalah bismillah hasil debat capres pertama mantap dgn paslon 1 tambah sudah ditandatanganinya pakta integritas makin yakin tidak 1 bukti pak anies orang amanah kalian bagaimana

Table 32. Stopwords Dataset Presidential Candidate Pabowo Subianto

Stemming
prabowo hampir pasti sulit kalah bahkan target paslon 02 level menang 1 putar makanya sangat wajar 01 03 lebih pilih serang 02 01 03 sulit juara 1 di putar 1 dua mungkin menang bisa masuk putar 2 reaksi dukung paslon 02 usia nobar debat capres sama sama aammiinkan moga allah swt tuhan milik kuasa kabul minta prabowo rakyat indonesia pilih paslon 02 aaammiin robb

Table 33. Stopwords Dataset Presidential Candidate Ganjar Pranowo

Stemming
yakin nonton besok rilis survey kalau capres 2 raih 97 menang telak paslon 1 2 capres 3 1 di survey 1 capres emosi 2 capres joget kl bisa jawab 3 capres endors jokowi mulu paslon 1 anies rasyid baswedan kalah penting hadir orang tua almarhum ananda harun al rasyid tengah capres 2 tidak 3 korban tinggal dunia tragedi 22 mei 2019 bapak din wahyudin harap adil usut tuntas next km 50 survei poltracking baru elektabilitas 3 paslon pipres 2024 prabowogibran 452 ganjarmahfud 273 aniesimin 231 ttj 44 paslon 01 03 sila kalian cakar2an rebut 1 tiket putar keduaklo paslon 02 kerja lebih keras menang 1 putar

Table 34. Stopwords Dataset Vice Presidential Candidate Muhaimin Iskandar

Stemming
mas gibran dulu pak muhaimin pernah ikut potong tumpeng ikn jadi konsisten gin yahhh kena tuh slepet cak imin el sulfat gus muhaimin aneh ken bangun kota level jakarta tidak tuju sama ikn ya monggo lah serah biarkan el sulfat masak cak imin paslon nomor urut 1 anies baswedan muhaimin iskandar cak imin sampai terang usai debat cawapres jumat 22122023 anies cak imin sempat komentar strategi pihak dalam beri tanya debat tahu debat cawapres

Table 35. Stopwords Dataset Vice Presidential Candidate Gibran Rakabuming

Stemming
sesi satu debat cawapres 1 amin cak imin slepetnya 710 2 prabowo gibran gibran gagas hilir digital 7510 3 gama pak mahfud proyeksi tumbuh 7 masa depan 7510 usai 2 debat pertama para caprescawapres siapa kirakira pilih overall nilai debat cawapres malam 1 cak imin 73 2 gibran 69 3 mahfud md 72 tuju engga ps gibran nilai rang ada prilaku songong sama orang tua

Table 36. Stopwords Dataset Vice Presidential Candidate Mahfud MD

Stemming
bicara soal hukum ekonomicocek banget kapasitas cawapres 3 prof jelas banyak pihak ekonomi kuat bangun atas pondasi hukum yang kuat sosok adalah prof mahfud md si pendekar hukum 3konomi nonton debatcaawapres yukk simak apa harap cawapres 3 prof beliau harap nya ssk di bentuk lpsk pasti hukum nya benar benar di tega ganjar mahfud satu set cawapres 3 prof mahfud md jangan minta buru jabatankarna sama kita kerja buat diri sendiribkn rakyat

f. Data Analysis

After going through the preprocessing stage, the next step is to label the sentiment class in the sentiment analysis of the presidential and vice presidential candidates. At this stage, the labeling process is carried out using training data from the TextBlob library and applying it to the Naive Bayes algorithm. This process involves utilizing training data to train the model so that the model can recognize sentiment patterns in the text. As a result, when the model is applied to data test, such as reviews related to presidential and vice presidential candidates, it can provide a sentiment classification, whether positive or negative.

1) $P(\text{Positif}|\text{Dokumen})$

$$= \frac{P(\text{Dokumen}|\text{Positif}) - P(\text{Positif})}{P(\text{Dokumen})}$$

2) $P(\text{Negatif}|\text{Dokumen})$

$$= \frac{P(\text{Dokumen}|\text{Negatif}) - P(\text{Negatif})}{P(\text{Dokumen})}$$

3) $P(\text{Netral}|\text{Dokumen})$

$$= \frac{P(\text{Dokumen}|\text{Netral}) - P(\text{Netral})}{P(\text{Dokumen})}$$

Based on the results of the Naïve Bayes analysis test, the results of public sentiment toward the Presidential and Vice Presidential Candidates can be shown in figure below.

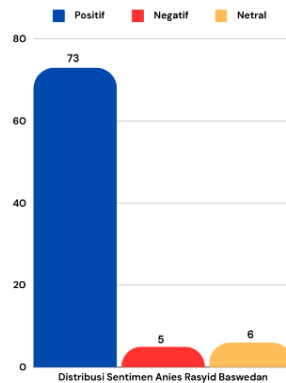


Figure 1. Public Sentiment Presidential Candidate Anies Baswedan

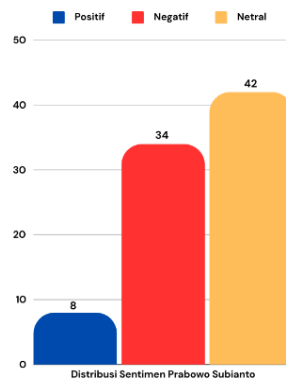


Figure 2. Public Sentiment Presidential Candidate Prabowo Subianto

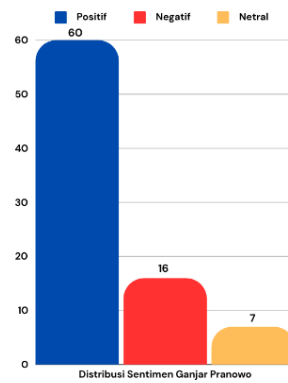


Figure 3. Public Sentiment Presidential Candidate Ganjar Pranowo

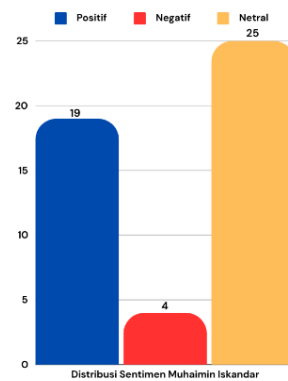


Figure 4. Public Sentiment Vice Presidential Candidate Muhaimin Iskandar

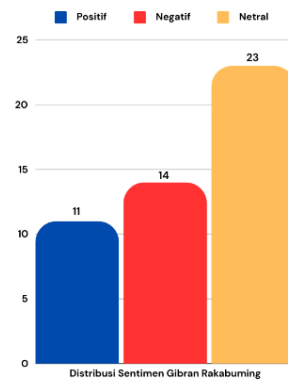


Figure 5. Public Sentiment Vice Presidential Candidate Gibran Rakabuming

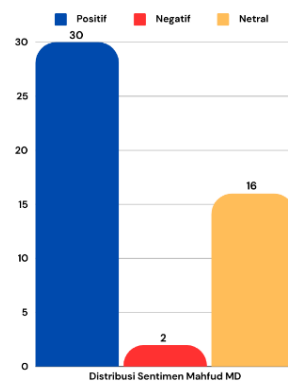


Figure 6. Public Sentiment Vice Presidential Candidate Mahfud MD

4. Discussion

From the analyzed tweet dataset, the result shows that the presidential candidate who received the most negative sentiments was Prabowo Subianto with 34 negative sentiments, 8 positive sentiments, and 42 neutral sentiments. Followed by candidate Ganjar Pranowo with 16 negative sentiments, 60 positive sentiments, and 7 neutral sentiments. Then the presidential candidate with the most positive sentiments was Anies Rasyid Baswedan with 73 positive sentiments, 5 negative sentiments, and 6 neutral sentiments. The vice presidential candidate who received the most positive sentiments was Mahfud MD with 30 positive sentiments, 2 negative sentiments, and 16 neutral sentiments. Followed by vice presidential candidate Muhaimin Iskandar with 19 positive sentiments, 4 negative sentiments, and 25 neutral sentiments. Then the vice presidential candidate with the lowest positive sentiments was Gibran with 11 positive sentiments, 14 negative sentiments, and 23 neutral sentiments.

4. Kesimpulan

This study explores the dynamics of public sentiment towards the Indonesian presidential and vice presidential candidates through tweet analysis on platform X. Involving the Naïve Bayes method, the results of the study provide an in-depth picture of the positive and negative responses exposed in digital discourse. Thus, this study makes a significant contribution to the understanding of public opinion in the digital era and the potential of social media as a mirror of public response to political figures. The implications of these findings can be used to design more effective campaign strategies and better understand the mindset and preferences of voters in the context of the presidential election. In addition, the methods applied in this study can be the basis for further research in political sentiment analysis on social media platforms.

Daftar Pustaka

- [1] K. Sistem and P. Di, "Jurnal Fakultas Hukum UNSRAT Lex Administratum Vol. 12 No. 4 Mei 2024," vol. 12, no. 4, 2024.
- [2] A. Nugroho, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, p. 200, 2018, doi: 10.30645/j-sakti.v2i2.83.
- [3] A. Fauzi, M. F. Akbar, and Y. F. A. Asmawan, "Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes," *J. Inform.*, vol. 6, no. 1, pp. 77–83, 2019, doi: 10.31311/ji.v6i1.5437.
- [4] A. Imron, "Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier," 2019.
- [5] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, doi: 10.33365/jatika.v3i2.2026.
- [6] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [7] E. Wittern *et al.*, "Opportunities in Software Engineering Research for Web API Consumption," *Proc. - 2017 IEEE/ACM 1st Int. Work. API Usage Evol. WAPI 2017*, pp. 7–10, 2017, doi: 10.1109/WAPI.2017.1.
- [8] E. Uzun, "A Novel Web Scraping Approach Using the Additional Information Obtained from Web Pages," *IEEE Access*, vol. 8, pp. 61726–61740, 2020, doi: 10.1109/ACCESS.2020.2984503.
- [9] F. Reza, C. K. Dewi, and E. Yudhyani, *Statistika Deskriptif untuk Ekonomi & Bisnis*. 2021.
- [10] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *J. Ilm. Comput. Sci.*, vol. 2, no. 2, pp. 89–98, 2024, doi: 10.58602/jics.v2i2.24.
- [11] S. W. S. S. 3) Muhamad Anton Permana1), "Analisis Sentimen Penggunaan Aplikasi VideoConference Pada Ulasan Google Play Store Menggunakan Metode Nbc (Naive Bayes Classifier)," *J. Ris. Sist. Inf. Dan Teknol. Inf.*, vol. 5, no. 1, pp. 178–191, 2023, [Online]. Available: <https://jursistekni.nusaputra.ac.id/article/view/178>.
- [12] R. G. Whendasmoro and J. Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 8
- [13] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.
- [14] O. Kaka, A. Ananda Putra Tanggu Mara, K. Wulla Rato, P. Studi Teknik Informatika, and S. Tinggi Manajemen Informatika Komputer Stella Maris Sumba Email Penulis Korespondensi, "Analisis

Sentimen Dampak Perkembangan Teknologi Informasi dan Komunikasi Terhadap Kemajuan Belajar Siswa SMK Rada Pamba dengan Metode Naive Baiyes,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, pp. 191–199, 2023.

- [15] Fauzan Baehaqi and N. Cahyono, “Analisis Sentimen Terhadap Cyberbullying Pada Komentar Di Instagram Menggunakan Algoritma Naïve Bayes,” *Indones. J. Comput. Sci.*, vol. 13, no. 1, pp. 1051–1063, 2024, doi: 10.33022/ijcs.v13i1.3301.