

Analyzing Abstention Discourse in Presidential Elections: Knowledge Discovery in X Using ML, LDA and SNA

Jayadi Butar Butar^a, Sofian Lusa^b, Sutia Handayani^a, Andi Akram Yusuf^a

^a Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

^b Institut Pariwisata Trisakti, Jakarta, Indonesia

Corresponding author: *jayadi.butar@ui.ac.id

Abstract—As a social media platform, X (formerly Twitter) has become a massive source of real-time, unstructured data, providing valuable insights into people's opinions on various issues. One crucial social phenomenon that has attracted attention on social media is the discourse around abstention (commonly known as “golput” in Indonesia) in the context of the presidential election. Abstention refers to the deliberate act of refusing to vote. Understanding the patterns, preferences, and topics associated with abstention discourse can provide valuable knowledge for political analysis. This study aims to discover knowledge based on patterns, sentiment polarization, and issues from unstructured X data to understand the discourse surrounding abstention in the 2024 Indonesian presidential election. The methodology involves collecting data from the X API, conducting Social Network Analysis (SNA) to analyze the social structure, preprocessing the data, and searching for the best sentiment analysis model through hyperparameter tuning on six Machine Learning (ML) models. Then, Latent Dirichlet Allocation (LDA) is employed with coherence score evaluation to identify topics related to the issue. The results indicate that 2,489 tweets discussing abstention were collected during the study period, exhibiting varied daily trends. SNA analysis reveals the formation of clusters within the dataset, alongside the identification of influential actors through three different centrality calculations. The sentiment analysis results show that the Logistic Regression (LR) model with count vectorizer is the best-performing model, with a predominance of positive sentiment polarity over negative sentiment. Evaluation of LDA using coherence scores indicates the presence of five topics related to abstention. This research contributes to knowledge discovery on the X platform by providing valuable insights into the discourse surrounding abstention in the Indonesian presidential election. These findings offer a deeper understanding of public opinion, political engagement, and election dynamics.

Keywords—Knowledge discovery; Twitter; abstention; topic modeling; social network analysis; machine learning.

*Manuscript received 16 Aug. 2023; revised 13 Mar. 2024; accepted 18 Apr. 2024. Date of publication 30 Jun. 2024.
IJASEIT is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.*



I. INTRODUCTION

The methodology of Knowledge Discovery (KD), which encompasses various activities such as data selection, preprocessing, transformation, and data mining, is pivotal in uncovering valuable patterns within data [1]. This systematic approach is crucial in tackling challenges posed by vast and unstructured datasets, such as those on social media platforms like X (formerly Twitter). With social media emerging as a rich real-time data source worldwide, the need to extract critical insights from this data has become increasingly prominent. Hence, the synergy between KD methodologies and the exploration of social media data underscores the growing importance of data mining and knowledge discovery in today's information-rich landscape.

Extracting critical information from social media data is one of the significant challenges in data mining and

knowledge discovery. Social media platforms like X have now become a vast source of unstructured and real-time data from across the globe [2]. This data encompasses a wide range of content, including tweets, comments, and blog posts, which contain valuable insights into people's opinions on various issues [3]. Governments, businesses, and individuals can harness this unstructured data to uncover valuable knowledge. Consequently, the exploration of such data has given rise to data mining and knowledge discovery. Knowledge discovery is the act of identifying patterns in data to be processed, with the output being beneficial information [4].

A significant portion of the data found on social media platforms can be attributed to the role of social media as a platform for individuals to express their opinions [5]. Every social phenomenon attracts people's attention to comment on social media, including political phenomena such as presidential election contests [6], [7]. The presidential

election is a phenomenon that is especially important for the political continuity of a country. That event determines the future direction of a country [8], [9]. Of course, every citizen has diverse political views manifest in their support for the presidential candidate, and their support is increasingly visible, especially one year before the presidential election. As voting day approaches, various opinions are expressed through social media, including support for a particular candidate, views on national issues, and discussions about abstention.

Abstention is an intentional act with a clear purpose of refusing to cast a vote in an election [10]. In Indonesia, abstention is commonly known as “*golput*” (*Golongan Putih*), referring to a form of moral movement to protest against the regime of Soeharto, who once held power in Indonesia [11]. In the 2014 Indonesian presidential election, 30.42% of registered voters did not exercise their right to vote. Furthermore, in the 2019 presidential election, 34.75 million people did not use their voting rights or participated in abstention. This number is equivalent to 18.02% of the Permanent Voter List (DPT) in the 2019 election, which consisted of 192.77 million people [11]. In Indonesia, no legal regulation prohibits abstention, but according to the election law, encouraging or advocating abstention is forbidden and can be considered a criminal act.

Recently, there has been a lot of research about elections using X to analyze public sentiment. One research project focuses on using social network analysis (SNA) to analyze X data [12]. The researchers have introduced a unique multilayer network structure with user and topic layers to gain meaningful insights from the network. The research evaluates different methods for identifying influential users and discovering communities, emphasizing the importance of sentiment analysis, topic modeling, and community detection in understanding user behavior and interactions on X. The findings suggest that by applying the SNA process, individuals and communities can be identified and targeted for specific marketing strategies. Another research on X data is about SNA toward X users against Hoax [13].

This research uses *Apache Hadoop Hortonworkstm* Distribution to gain the best result for SNA. This proves that X is still a hot topic for public opinion. Chakraborty and Mukherjee [14] discusses a study on analyzing and mining an election-based network using large-scale X data. The study explores the structure and dynamics of the state assembly election-based tweet-reply network generated by X users across India for 6 weeks. The authors use SNA and mining techniques to identify hashtags used by political contenders, cluster-level dominance in the X network, and network-level centrality measures to obtain in-depth inferences about the behavior and role of different network actors. The findings reveal fascinating insights into the flow of X activity and how this information can be used as a forecasting tool. Another research [10] analyzed discourses related to voter abstention during the 2017 French presidential election, specifically focusing on X hashtags. This research aims to gain insights into public discussions about abstention in the modern digital era. The study found that discussions surrounding abstention revealed a significant level of distrust in the current state of French democracy. These discussions also raised concerns about voter manipulation and expressed opposition towards

the incoming president. The results suggest that conversations about trust in French democracy, especially among specific groups, are concerning. On X, abstaining from voting was viewed as an “active” way of protesting against a perceived corrupt and manipulated political system. Another study addressed the influence of social media influencers on the outcome of the 2019 election in Indonesia, focusing on X conversational activity around presidential candidates [15]. This research uses SNA to measure public conversation activity through social media networks and machine learning (ML) to analyze the sentiment polarity of public conversations. The study shows that using SNA and sentiment analysis can prove that influencers can predict election results through their expressions on X. Thus, a relationship exists between influencers' expressions on social media and the winning party in the 2019 Indonesian election.

Previous studies have shown that X data is a common source of information used to collect topic extraction and public sentiment and find influential figures in public opinion. Some commonly used methods are SNA, Sentiment Analysis with ML, and LDA for topic modeling. In this research, SNA is used to find critical factors in abstention-related topics, sentiment analysis with ML is used to see the polarity of public opinion related to abstention, and LDA is used to find the most discussed issues related to abstention. Abstention is a fascinating topic in an election. Especially in Indonesia's Presidential Election, as the population is vast and very diverse, there are never more than three candidates. This means that three candidates may not be enough to represent the 273.8 million people of Indonesia [16].

The research problem addressed in this study is discovering knowledge based on patterns, sentiment polarization, and topics from unstructured X data to understand the discourse surrounding abstention in the 2024 Indonesian presidential election. Several studies have been conducted in the electoral field, including predicting elections using X to predict the results using various methods [17], [18]. The underlying logic of these studies is to gather data using the X API, apply different classification techniques, and identify distinct trends among them to predict the election [19]. However, few studies have specifically analyzed the abstention discourse on X. Therefore, we use SNA to identify the key players in the abstention discourse on X, Sentiment analysis with several machine learning algorithms to determine the sentiment polarity of the tweets, and Latent Dirichlet Allocation (LDA) to determine the related topics of the abstention discourse shared on X.

II. MATERIALS AND METHOD

Fig. 1 illustrates the proposed framework steps in this research. As indicated in the figure, social network analysis (SNA) is conducted after collecting the dataset, which occurs before data preprocessing, as SNA does not require preprocessed data. After that, we proceed to the preprocessing step, which includes various data cleaning techniques, text normalization, and stop word removal. Following preprocessing, features are selected and extracted using a combination of count vectorizer and TF-IDF approaches to generate vectors. Then, a series of ML classification algorithms with hyperparameter tuning are applied to classify tweets as positive or negative. The classification results are

evaluated by comparing their F1 scores. Finally, a topic modeling process is conducted to discover related topics.

These steps are elaborated in detail in the following subsections.

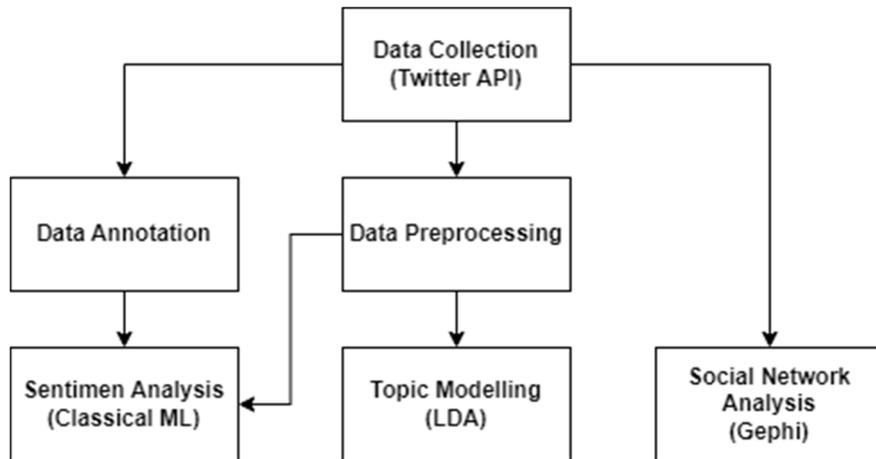


Fig. 1 Research methodology.

A. Data Collection

The first stage is data collection, where data is gathered using the X API, resulting in data in CSV format [20], [14]. The dataset was collected randomly from May 13 to 31, 2023, using the hashtag #golput, widely used in tweets containing discourse on abstention.

B. Social Network Analysis

Social network analysis (SNA) is an analysis process of the interaction between humans, investigating the social structure within a group of people [21]. SNA is a popular method that has been used many times. SNA can process a large amount of data and describe the relationships between each node; most of the time, the node represents a human. Notably, SNA can represent the data in a graph/diagram model that is easier to understand. This diagram can also be a means to convey the result of an analysis. SNA is commonly paired with X data because X data can have a mention or retweet, which can be implemented into SNA. In this case, the node represents an account [22].

The collected X data is analyzed using network and graph theory, and the SNA method is applied to examine the social structure. This method classifies the network system into nodes (in this case, X accounts) and ties, edges, or links (such as retweets and comments) that connect them. The tool used in this stage is the Gephi application, an open-source software for visualizing and exploring various graphs and networks. The outcome of this stage is the identification of Key Actors who play a role in disseminating the discourse on abstention [23].

C. Annotation

The annotation process is required for the sentiment analysis stage. The annotation process is done by manually determining the sentiment polarity. polarity consists of two nominal values: positive and negative.

D. Preprocessing

Data sourced from X, primarily comprising public sentiments about elections, is textual. Nonetheless, this information frequently encompasses extraneous elements,

which can obfuscate subsequent analysis [24]. Consequently, it is imperative to execute data pre-processing to eliminate irrelevant terms from these tweets [25], [26]. This stage consists of sub-stages as follows:

- Text cleaning: Eliminating non-alphanumeric characters from the text using regular expressions.
- Character conversion: Converting emoticons, URLs, and mentions.
- Case-folding: converting sentences to lowercase.
- Tokenization: Select all words that appear in tweets, remove symbols and punctuation marks in each tweet, and treat them as separate tokens.
- Stop word Removal: cleaning the text from words belonging to stop words, commonly used words that do not have significant meaning.
- Normalization: transforming the text or words into a standardized form, such as converting abbreviations or variations of words into their complete forms.

E. Sentiment Analysis

Sentiment analysis is a method for identifying opinions and sentiments in text. It consists of detecting, extracting, and classifying views on a subject, which involves Natural Language Processing (NLP) to track public views on a particular topic [27]. Sentiment analysis generally means studying people's sentiments towards a specific entity.

In this study, a supervised machine learning approach is employed utilizing six algorithms, namely Support Vector Machine (SVM) [28], Logistic Regression (LR) [29], Naive Bayes (NB) [30], Decision Tree (DT) [31], Random Forest (RF) [32], and Extreme Gradient Boosting (XGBoost) [33]. Before building and comparing sentiment analysis models, hyperparameter tuning [33], [3] for the 6 algorithms using grid search and 5-fold cross-validation techniques was performed. The purpose of the following process is to identify the optimal parameter values. These parameter values are not learned from the data but are determined before the model construction is run. For example, in the case of Logistic Regression (LR), parameters such as regularization strength (C) and penalty (l1 or l2) are tuned to find the settings with the best classification performance. Table 1 shows the results

of the hyperparameter tuning, which ultimately resulted in the best model version for each machine learning algorithm.

TABLE I
HYPERPARAMETER TUNING

Model	Hyperparameters	Accuracy
SVM	C: 100; gamma: 0.001; kernel: rbf.	0.78
LR	C: 1; penalty: l2; solver: newton-cg.	0.77
RF	bootstrap: False; max_depth: 20; min_samples_leaf: 1; min_samples_split: 10; n_estimators: 300.	0.78
DT	criterion: gin; max_depth: None; min_samples_leaf: 1; min_samples_split: 10; splitter: random.	0.77
NB	alpha: 0.1.	0.68
XGB	colsample_bytree: 0.5; gamma: 0.3; learning_rate: 0.1; max_depth: 5; min_child_weight: 1; subsample: 0.9.	0.76

Two feature engineering approaches, namely TF-IDF and Count Vectorizer, are used. TF-IDF is a matrix representation of the dataset, where each row represents a document in the corpus, each column represents a token in the corpus, and the value in each cell indicates the number of occurrences of the feature in each document in the corpus [34]. These values change inversely proportional to the frequency of occurrence of the feature; in other words, the more frequently a feature appears, the smaller its TF-IDF value. TF-IDF is a very commonly used vectorization technique, which often gives better results in accuracy and can be applied to both unigrams and n-grams. Meanwhile, Count Vectorizer has a similar structure to TF-IDF in terms of rows and columns, but the values in the cells represent the frequency of occurrence of a token in a document without taking into account the occurrence of that token in other documents [33].

F. Topic Modeling

Topic modeling is one of the few methods in text mining to identify a topic from a text document and one of the most popular methods is Latent Dirichlet Allocation (LDA). LDA is a statistical model representing the topic of discussion in a text [35]. The basic concept of LDA is an article must have a topic, some even have a multiple topic, and machine need a statistical model to understand the topic of a text, thus leading to the surfacing of LDA. The concept is that if a text has a cat-related topic, it must have multiple cat words, such as kitten, milk, meow, etc. LDA needs clean data before processing a document; the clean data means the sentence did not have a stop word, such as “is,” “the,” and “and.” Based on this concept, LDA will generate a topic for each document. LDA is a widespread topic extraction, proved by an existing library for LDA [36]. Another notable is that there has been a lot of previous research using LDA.

To get the best number of topics in LDA, evaluation by coherence score is used [33]. For this purpose, several LDA models with different numbers of topics are created and compared based on the coherence score to find the best number of topics. In this research, LDA is implemented using the *Gensim* library. The result of this stage is identifying issues related to the abstention discourse.

III. RESULTS AND DISCUSSION

The research results indicate that from May 13, 2023, to

May 31, 2023, there were 2489 tweets discussing abstention, including tweets, quotes, retweets, and replies. Figure 2 shows the number of tweets varied each day. The trend peaked on May 14, 2023, with 468 tweets before declining from May 15, 2023 (263 tweets) to May 18, 2023 (105 tweets). On May 19, the trend picked up again and reached 328 tweets on May 21. However, the trend declined to 7 tweets by May 31. Meanwhile, Figure 3 shows the trends of all types of tweets dominated by replies.

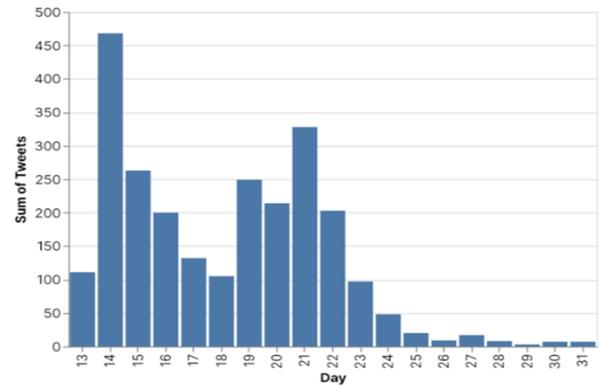


Fig. 2 Total tweets

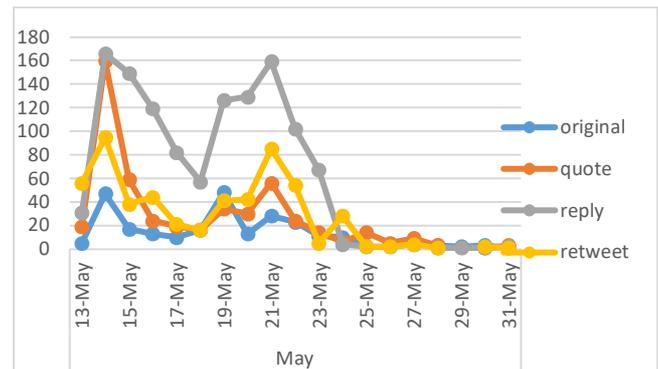


Fig. 3 Trends of total tweets

A. Key Players

Figure 4 depicts the interrelationships among actors in the discourse of abstention on X. In this graph, each node represents a user, and the edges indicate interactions between users through various tweet actions, such as retweeting, quoting, or replying to another user's tweet. The graph illustrates three major clusters. The first cluster is shown in pink, representing a large network with user id 1640775351826800xxx as the main actor.

The second cluster is green, and the user with ID 1533832993877291xxx plays a central role in its dynamics. The third cluster is blue, with the main actor being an account with user ID 1603676333527076xxx. In addition to the influential actors within each cluster, numerous actors (nodes) are not connected to the larger clusters. This network exhibits specific characteristics based on its statistical properties. The average degree value is 0.708, indicating that, on average, each user interacts with approximately 0.708 other users through tweet actions like retweets, quotes, or replies. The average clustering coefficient is relatively low at 0.006, suggesting that there are few connections between users'

immediate neighbors, making the occurrence of a "small world" network effect [37] unlikely occur.

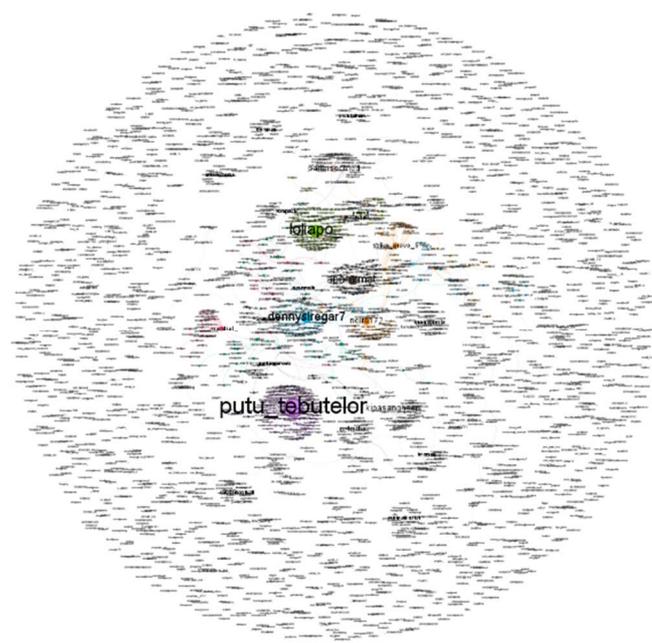


Fig. 4 Network clusters of actors in X

The network's structure resembles concentric rings, where highly connected nodes cluster at the ring's core. This configuration results from limited connectivity between most nodes, leading to a longer propagation time for information to spread among connected users. The low density of the network, calculated as 0, reinforces the sparse connections between nodes. Moreover, the network's diameter, determined as 4, indicates the maximum number of steps required to reach the furthest pair of nodes. This further emphasizes the limited connectivity between users and the time it takes for information to traverse the network. This network forms the Community Clusters Characteristics [38], where the conversation pattern resembles a bazaar with different stalls characterized by several even-sized groups rather than a crowd of primarily unconnected nodes.

Table II provides the top 5 nodes with the highest degree values, arranged in descending order of their degrees. These nodes are the most highly connected users in the network.

TABLE II
TOP DEGREE

No	User ID	Degree
1	1640775351826800xxx	111
2	1533832993877291xxx	67
3	1603676333527076xxx	51
4	40043xxx	47
5	1127334227253481xxx	33

To identify key players in SNA, the parameters called centrality is used [39], [40]. Centrality measures the importance of actors using metrics in a social network. This study employed two centrality metrics to identify influential users on X: Betweenness Centrality, and Closeness Centrality.

TABLE III
BETWEENNESS AND CLOSENESS CENTRALITY

Betweenness Centrality		Closeness Centrality	
User ID	Value	User ID	Value
183332xxx	0.04054	115438xxx	0.8
725349607807455xxx	0.037371	1376387921960308xxx	0.75
89418xxx	0.037033	1376387921960308xxx	0.75
1252247274006044xx	0.030794	1487742160237568xxx	0.727
x			
1603676333527076xx	0.029809	1533832993877291xxx	0.7
x			

Betweenness Centrality is derived from network theory and has been previously used in similar research [39]. This metric aims to identify users who significantly influence others through their tweets. These influential users are often strategically positioned within the network and play a role in information propagation. The scores obtained from betweenness centrality can be used to rank users from the most influential to the least influential, considering the importance of nodes in connecting other nodes. The graph's betweenness centrality calculation results show that the highest value betweenness centrality is an X user ID 183332xxx, with a score of 0.04054.

Closeness Centrality is used to measure how closely connected an actor is to all other actors in a social network [41]. This centrality measure is essential for determining the expenditure of influential actors within the network. Closeness is measured by the number of steps or paths it takes for an actor to reach or be reached by other actors in the network. Closeness centrality has a variant called harmonic closeness centrality that was invented to solve the original formula's problem when dealing with unconnected graphs. The result from the calculation shows that the highest value closeness centrality is an X account 115438xxx, a score of 0.8.

B. Sentiment Analysis

Table IV evaluates the performance of the six models created using three different vectorization methods and 5-fold cross-validation. It compares models based on F1-score performance metrics.

TABLE IV
MODEL PERFORMANCE EVALUATION

Vectorizer	DT	LR	NB	RF	SVM	Xgb
Count Vectorizer	0.90	0.94	0.83	0.88	0.93	0.90
TF-IDF Unigram	0.89	0.90	0.82	0.87	0.88	0.89
TF-IDF Bigram	0.83	0.84	0.83	0.77	0.77	0.79
TF-IDF Trigram	0.78	0.78	0.79	0.70	0.70	0.71

The model evaluation indicates that the logistic regression model achieved the highest F1-Score of 0.94 using a count vectorizer. Overall, the count vectorizer outperforms Tf-idf unigram, bigram, and trigram for all models. Additionally, LR also demonstrates superior performance compared to other models.

Figure 5 depicts the sentiment analysis results for all X accounts discussing abstention, revealing that CybernetWalker predominantly discusses abstention with negative and minimal positive sentiment. In contrast, account NkriRindu predominantly discusses abstention with positive sentiment. The sentiment analysis results also indicate that the polarity of discussions on abstention on X leans more towards positive sentiment, with the number not significantly differing

from pessimistic sentiment. This suggests that many accounts discuss this topic with a positive narrative, albeit with fewer discussions. In contrast, fewer accounts discuss this negatively but with more intense discussion intensity.

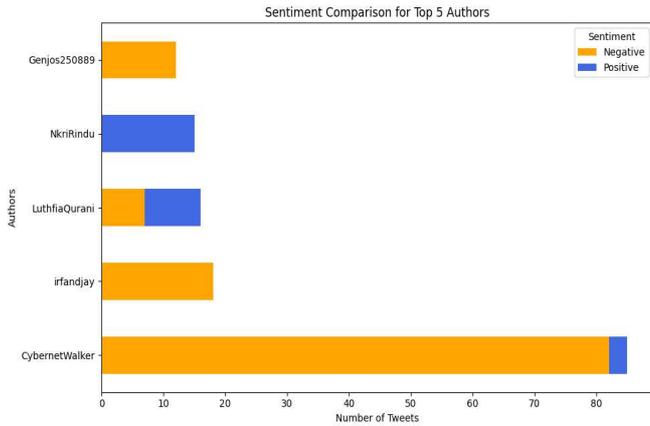


Fig. 5 Sentiment Analysis result

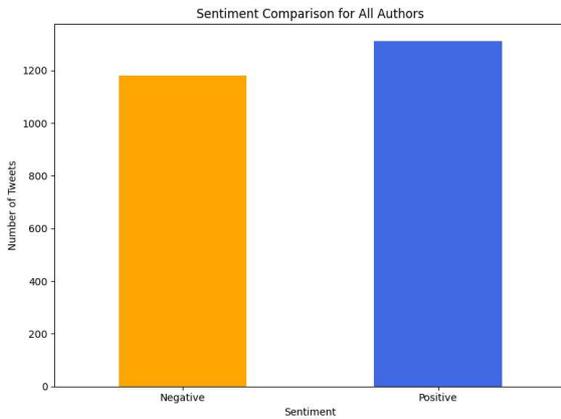


Fig. 6 Sentiment Polarization

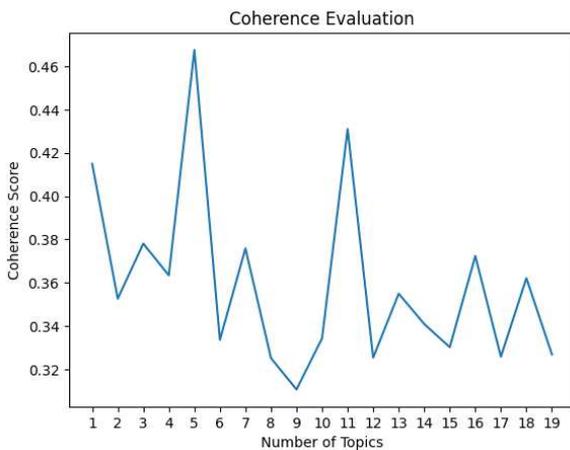


Fig. 7 Coherence score

C. Topic Extraction

Before creating LDA, a search for the optimal number of topics was conducted using a coherence score. Figure 7 displays the coherence score graph, indicating that five topics, with a coherence score of 0.47, are the most suitable for this dataset.

Table V summarizes the top 5 topics among all the identified topics. Table 3 shows that the top 6 words represent a subtopic related to voter abstention (*golput*). By analyzing these words, we can categorize the topics into discussions pertaining to voter abstention.

TABLE V
TOPIC EXTRACTION

No.	Topic	Description	Significant Word
1	Democracy	The belief that abstention is also a part of a democratic attitude.	Demokrasi, suara, coblos, pilih, golput, bersih
2	Candidates	public's dissatisfaction with the presidential candidates which are being widely discussed	Mending, prabowo, capres, ganjar, pilih, golput
3	Right to vote.	An encouragement to exercise the right to vote and avoid abstention.	Suara, cerdas, masyarakat, twitter, harap, hitung
4	Ballot papers	Concerns about unused ballot papers being misused.	Suara, golput, coblos, kertas, jahat, pakai
5	Change	Reasons for voters not to vote because they feel there will be no change.	Golput, coblos, kuasa, orba, muak, tetap

Topic #1 is the belief that abstaining is also part of a democratic attitude. Topics close to the most prominent words are 'democracy,' 'votes,' 'choose,' 'vote,' 'abstentions,' and 'clean', which describe why people abstain.

Topic #2 contains the reasons why people choose to abstention. This topic is close to the most prominent words: 'mending', 'pilih', and 'golput'. It also contains the names of the most discussed presidential candidates, such as Prabowo and Ganjar. This indicates the public's dissatisfaction with the presidential candidates being widely discussed.

Topic #3 is a social media campaign inviting people not to abstain from voting. This topic contains prominent words such as 'suara', 'cerdas', 'masyarakat', 'twitter', 'harap', 'hitung'. This topic is closely related to the Smart Voting Movement, which election organizers (KPU and BAWASLU) actively promote.

Topic #4 contains the concerns of some people about the misuse of unused ballots when many people vote to abstain. This topic is related to the words 'suara', 'golput', 'coblos', 'kertas', 'jahat', 'pakai'.

Topic #5 is related to why people abstain. words like 'golput', 'coblos', 'kuasa', 'orba', 'muak', 'tetap' are related to this topic. Specifically, the word 'orba' is a disappointment for people who feel that there has been no change in their lives since the New Order era (the reign of President Soeharto, which ended in 1998).

IV. CONCLUSION

This study focuses on discovering knowledge based on patterns, preferences, and topics from unstructured X data to understand the discourse surrounding abstention (*golput*) on the X platform in Indonesia. The discovered knowledge was analyzed using Social Network Analysis (SNA), supervised

Machine Learning, and Latent Dirichlet Allocation (LDA) methods. The Results of our study lead to several conclusions.

First, SNA reveals the existence of three major clusters surrounded by several smaller clusters in the discourse on abstention on X. In addition to cluster formation, some users have the most significant influence and are considered critical actors in spreading the issue. These users are identified as "1640775351826800xxx" with the highest degree level or the center of the largest cluster, meaning that this issue is most discussed on this account. Furthermore, the "183332xxx" account serves as a bridge between clusters in the network (highest betweenness centrality), indicating its activity in discussing the issue in posts on other accounts. Meanwhile, the "115438xxx" user-id has the highest closeness centrality, suggesting that this account has many friends who often discuss the issue of abstention. Characteristics of the network community clusters from this analysis indicate that several influencers dominate the discourse on this issue on X, and there is little discussion or conversation between users and groups. This finding may be related to X being widely used by individuals who may be talking about the same thing [42].

Second, In the sentiment analysis phase, it was found that using a count vectorizer consistently yielded better performance than the TF-IDF unigram, bigram, and trigram methods for all evaluated models. The evaluation showed that the logistic regression (LR) model achieved the highest F1-Score of 0.94 using count vectorizer, confirming its superiority in modeling the dataset. Sentiment analysis of conversations on X about abstention revealed a dominance of positive sentiment, which is surprising considering people's tendency to give more weight to negative aspects and the supremacy of negativity in political discourse on social media [43]. However, this finding may be influenced by the presence of influencers intentionally encouraging people not to abstain from voting, similar to the findings of previous research [44] where influencer accounts tended to be more active in sharing positively sentimental political content. The high intensity of discussion with predominantly negative sentiment from a few accounts suggests the possibility of 'buzzers' deliberately spreading abstention issues.

Third, based on coherence score evaluation in the topic modeling process, the dataset was best grouped into 5 topics, each represented by 6 dominant words. The resulting issues include the public belief that abstention is also part of a democratic attitude, public dissatisfaction with extensively discussed presidential candidates, campaigns against abstention, concerns about misusing unused ballot papers, and public disappointment with the perceived lack of change. This finding is intriguing because there are more negative topics than positive ones, despite the higher number of positive sentiment tweets compared to negative sentiment tweets. This indicates that although there is more discussion on abstention with positive sentiment on X, the topics discussed are mainly related to campaigns urging people not to abstain from voting. Meanwhile, discussions with negative sentiment, although fewer, cover a broader range of topics.

In summary, the research findings indicate that patterns of conversation regarding abstention can be detected from X datasets, providing insights into the political landscape in Indonesia leading up to the 2024 presidential election. Conversations about abstention are polarized, reflecting

diverse perspectives on the topic and revealing the distribution of opinions from key figures. Additionally, specific accounts actively discuss abstention with negative sentiments on the X social media platform. Relevant parties can utilize this knowledge to develop strategies to increase voter participation in upcoming elections, map X accounts that promote abstention, evaluate the reasons behind individuals' decisions not to vote, and more.

During the research process, several limitations were encountered that could guide future research. Firstly, data collection on X was conducted using a free developer account owned by the author, resulting in limited access to past tweet data on X. Additionally, the narrow duration of data collection may lead to discussions reflecting only the public responses within that timeframe. The dynamic political situation leading up to the elections can result in rapid changes in views on political issues such as abstention; future research could use time range analysis to better explain the dynamics of this issue. Secondly, due to the focus on text data on the X social media platform, this study did not encompass the public's views from other platforms such as Facebook and YouTube and other forms of data (e.g., photos, videos, memes, etc.). This could lead to interpretations of research results that do not accurately represent the general public's views. Future research has the potential to include other forms of data and social media platforms to ensure that the results better represent the broader views of the public.

Finally, according to X's guidelines on using X user data for research, this study was conducted with careful attention to ethical aspects of X research to ensure that users are protected, and their rights are respected. User confidentiality is maintained by masking user identities, and in the data collection stage, the research follows the Decision flow chart for the publication of X communications [45].

ACKNOWLEDGMENT

This research is supported by the Indonesia Endowment Fund for Education (LPDP-Indonesia).

REFERENCES

- [1] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Soc. Sci. Res.*, vol. 110, no. April 2022, p. 102817, 2023, doi: 10.1016/j.ssresearch.2022.102817.
- [2] C. Zhang and J. Han, "Data Mining and Knowledge Discovery," *The Urban Book Series*, pp. 797–814, 2021, doi: 10.1007/978-981-15-8983-6_42.
- [3] M. Aljabri, S. S. Aljameel, I. U. Khan, N. Aslam, S. M. B. Chrouf, and N. Alzahrani, "Machine Learning Model for Sentiment Analysis of COVID-19 Tweets," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 12, no. 3, pp. 1206–1214, 2022, doi: 10.18517/ijaseit.12.3.14724.
- [4] K. L. Tsui, V. Chen, W. Jiang, F. Yang, and C. Kan, "Data Mining Methods and Applications," *Springer Handbooks*, pp. 797–816, 2023, doi: 10.1007/978-1-4471-7503-2_38.
- [5] N. Pokhriyal, B. A. Valentino, and S. Vosoughi, "Quantifying participation biases on social media," *EPJ Data Sci.*, vol. 12, no. 1, 2023, doi: 10.1140/epjds/s13688-023-00405-6.
- [6] J. J. E. Macrohon, C. N. Villavicencio, X. A. Inbaraj, and J.-H. Jeng, "A Semi-Supervised Approach to Sentiment Analysis of Tweets during the 2022 Philippine Presidential Election," *Inf.*, vol. 13, no. 10, 2022, doi: 10.3390/info13100484.
- [7] S. Limboi and L. Diosan, "An unsupervised approach for Twitter Sentiment Analysis of USA 2020 Presidential Election," *16th Int. Conf. Innov. Intell. Syst. Appl. INISTA 2022*, pp. 1–6, 2022, doi:10.1109/INISTA55318.2022.9894264.

- [8] T. L. S. Pinto, B. M. Tabak, and D. O. Cajueiro, "How politics can influence the allocation of social program benefits: A case study of the Brazilian poverty reduction program Bolsa Familia," *Econ. Anal. Policy*, vol. 80, pp. 77–89, 2023, doi: 10.1016/j.eap.2023.07.009.
- [9] A. Vigna-Gómez, J. Murillo, M. Ramirez, A. Borbolla, I. Márquez, and P. K. Ray, "Design and analysis of tweet-based election models for the 2021 Mexican legislative election," *EPJ Data Sci.*, vol. 12, no. 1, 2023, doi: 10.1140/epjds/s13688-023-00401-w.
- [10] J. Downing and E. E. Brun, "'I Think Therefore I Don't Vote': discourses on abstention, distrust and twitter politics in the 2017 French presidential election," *French Polit.*, vol. 20, no. 2, pp. 147–166, 2022, doi: 10.1057/s41253-021-00166-6.
- [11] T. Duile, "Challenging Hegemony: Nurhadi-Aldo and the 2019 Election in Indonesia," *J. Contemp. Asia*, vol. 51, no. 4, pp. 537–563, 2021, doi: 10.1080/00472336.2020.1748896.
- [12] A. P. Logan, P. M. LaCasse, and B. J. Lunday, "Social network analysis of Twitter interactions: a directed multilayer network approach," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, pp. 1–18, 2023, doi:10.1007/s13278-023-01063-2.
- [13] H. F. Karimi, Arini, S. U. Masrurroh, and F. Mintarsih, "The Influence of Iteration Calculation Manipulation on Social Network Analysis Toward Twitter's Users Against Hoax in Indonesia with Single Cluster Multi-Node Method Using Apache Hadoop Hortonworkstm Distribution," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–6, 2019, doi:10.1109/CITSM.2018.8674374.
- [14] A. Chakraborty and N. Mukherjee, "Analysis and mining of an election-based network using large-scale twitter data: a retrospective study," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, pp. 1–20, 2023, doi:10.1007/s13278-023-01081-0.
- [15] O. Monica, F. W. Wahida, and H. Fakhruroja, "The Relations between Influencers in Social Media and the Election Winning Party 2019," *Proceeding - 2019 Int. Conf. ICT Smart Soc. Innov. Transform. Towar. Smart Reg. ICISS 2019*, pp. 0–4, 2019, doi:10.1109/ICISS48059.2019.8969801.
- [16] J. G. O'Reilly, "A proposal to strengthen Indonesian democracy," *Asian Polit. Policy*, p. 12705, 2023, doi: 10.1111/aspp.12705.
- [17] M. Mohd et al., "poliWeet — Election prediction tool using tweets," *Software Impacts*, vol. 17, p. 100542, Sep. 2023, doi:10.1016/j.simpa.2023.100542.
- [18] P. Chauhan, N. Sharma, and G. Sikka, "Application of Twitter sentiment analysis in election prediction: a case study of 2019 Indian general election," *Social Network Analysis and Mining*, vol. 13, no. 1, May 2023, doi: 10.1007/s13278-023-01087-8.
- [19] D. Kumar and F. Ahamad, "Opinion Extraction from Big Social Data Using Machine Learning Techniques: A Survey," in *AIP Conference Proceedings*, 2023, vol. 2916, no. 1, doi: 10.1063/5.0179023.
- [20] H. Ali, H. Farman, H. Yar, Z. Khan, S. Habib, and A. Ammar, "Deep learning-based election results prediction using Twitter activity," *Soft Comput.*, vol. 26, no. 16, pp. 7535–7543, 2022, doi: 10.1007/s00500-021-06569-5.
- [21] Z. Dai and C. Higgs, "Social Network and Semantic Analysis of Roe v. Wade's Reversal on Twitter," *Soc. Sci. Comput. Rev.*, no. 0123456789, 2023, doi: 10.1177/08944393231178602.
- [22] A. P. Logan, P. M. LaCasse, and B. J. Lunday, "Social network analysis of Twitter interactions: a directed multilayer network approach," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, 2023, doi:10.1007/s13278-023-01063-2.
- [23] N. A. Diyana Suhaimi, S. Iffah Mohammad Salleh, S. A. Farhanah Abdul Hakim, P. Magalingam, N. B. Maarop, and M. Shanmugam, "Malaysian Politicians' Connection Pattern on Twitter using SNA: A Case of Najib Razak," *2021 Int. Congr. Adv. Technol. Eng. ICOTEN 2021*, no. May 2018, pp. 1–10, 2021, doi:10.1109/ICOTEN52080.2021.9493501.
- [24] T. K. V. Sai, K. Lohith, M. P. Sai, K. Tejaswi, P. M. Ashok Kumar, and C. Karthikeyan, "Text Analysis on Twitter Data Using LSA and LDA," *2023 Int. Conf. Comput. Commun. Informatics, ICCCI 2023*, no. Iccci, pp. 1–6, 2023, doi: 10.1109/ICCCI56745.2023.10128417.
- [25] V. C. Hardita, R. Hammad, and A. Z. Amrullah, "Mandalika Modeling Topic on Social Media Using Latent Dirichlet Allocation," *ICCoSITE 2023 - Int. Conf. Comput. Sci. Inf. Technol. Eng. Digit. Transform. Strateg. Facing VUCA TUNA Era*, pp. 1–5, 2023, doi:10.1109/ICCoSITE57641.2023.10127821.
- [26] K. Sethia, M. Saxena, M. Goyal, and R. K. Yadav, "Framework for Topic Modeling using BERT, LDA and K-Means," *2022 2nd Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2022*, pp. 2204–2208, 2022, doi: 10.1109/ICACITE53722.2022.9823442.
- [27] E. Del Valle Martín and L. De La Fuente Valentin, "Sentiment analysis methods for politics and hate speech contents in Spanish language: a systematic review," *IEEE Lat. Am. Trans.*, vol. 21, no. 3, pp. 408–418, 2023, doi: 10.1109/TLA.2023.10068844.
- [28] M. Wyawahare, A. Dhanawade, M. Dhopade, S. Dharyekar, and A. Dhole, "Twitter sentiment analysis on political tweets," *AIP Conf. Proc.*, vol. 2981, no. 1, p. 20022, Dec. 2023, doi: 10.1063/5.0182743.
- [29] A. A. Lina, H. Khadidja, J. Imene, and M. Neila, "Analyzing US Airline Customer Sentiment on Twitter using Multinomial Logistic Regression and Feature Reduction," *Colloq. Inf. Sci. Technol. Cist*, pp. 265–270, 2023, doi: 10.1109/CiSt56084.2023.10409979.
- [30] P. Assiroj, A. Kurnia, and S. Alam, "The performance of Naïve Bayes, support vector machine, and logistic regression on Indonesia immigration sentiment analysis," *Bull. Electr. Eng. Informatics*, vol. 12, no. 6, pp. 3843–3852, 2023, doi: 10.11591/eei.v12i6.5688.
- [31] R. Chen and R. Dong, "The Relationship Between Twitter Sentiment and Stock Performance: A Decision Tree Approach," *Proceedings of the 56th Hawaii International Conference on System Sciences*, 2023, doi: 10.24251/hicss.2023.592.
- [32] G. Thangarasu and K. R. Alla, "Detection of Cyberbullying Tweets in Twitter Media Using Random Forest Classification," *13th IEEE Symp. Comput. Appl. Ind. Electron. ISCAIE 2023*, pp. 113–117, 2023, doi: 10.1109/ISCAIE57739.2023.10165118.
- [33] E. Edgari, J. Thiojaya, and N. N. Qomariyah, "The Impact of Twitter Sentiment Analysis on Bitcoin Price during COVID-19 with XGBoost," *5th Int. Conf. Comput. Informatics, ICCI 2022*, pp. 337–342, 2022, doi: 10.1109/ICCI54321.2022.9756123.
- [34] Y. Setiawan, N. Maulidevi, and K. Surendro, "The Use of Dynamic n-Gram to Enhance TF-IDF Features Extraction for Bahasa Indonesia Cyberbullying Classification," 2023, pp. 200–205, doi:10.1145/3587828.3587858.
- [35] K. Kukushkin, Y. Ryabov, and A. Borovkov, "Digital Twins: A Systematic Literature Review Based on Data Analysis and Topic Modeling," *Data*, vol. 7, no. 12, 2022, doi: 10.3390/data7120173.
- [36] B. V. Barde and A. M. Bainwad, "An overview of topic modeling methods and tools," *Proc. 2017 Int. Conf. Intell. Comput. Control Syst. ICICCS 2017*, vol. 2018-Janua, pp. 745–750, 2017, doi:10.1109/ICCONS.2017.8250563.
- [37] M. Tanhapour, A. A. Safaei, and H. Shakibian, "Personal health record system based on social network analysis," *Multimedia Tools and Applications*, vol. 81, no. 19, pp. 27601–27628, Mar. 2022, doi:10.1007/s11042-022-12910-3.
- [38] A. Gerber, "The Detection of Conversation Patterns in South African Political Tweets Through Social Network Analysis," *Commun. Comput. Inf. Sci.*, vol. 1551 CCIS, pp. 15–31, 2022, doi:10.1007/978-3-030-95070-5_2.
- [39] A. J. Oroh, Y. Bandung, and L. M. Zagi, "Detection of the Key Actor of Issues Spreading Based on Social Network Analysis in Twitter Social Media," *Proc. - 2021 IEEE Asia Pacific Conf. Wirel. Mobile, APWiMob 2021*, pp. 206–212, 2021, doi:10.1109/APWiMob51111.2021.9435268.
- [40] R. Parvathi, Y. S. Asish, and V. Pattabiraman, "Analysis report for statistics in the Twitter network," *Deep Nat. Lang. Process. AI Appl. Ind. 5.0*, pp. 50–58, 2021, doi: 10.4018/978-1-7998-7728-8.ch003.
- [41] I. Williams, "Contemporary Applications of Actor Network Theory," *Contemp. Appl. Actor Netw. Theory*, pp. 1–283, 2020, doi:10.1007/978-981-15-7066-7.
- [42] J. Forestal, "Social Media, Social Control, and the Politics of Public Shaming," *Am. Polit. Sci. Rev.*, pp. 1–15, 2023, doi:10.1017/S0003055423001053.
- [43] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science (80-)*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: 10.1126/science.aap9559.
- [44] J. Weismueller, P. Harrigan, K. Coussemant, and T. Tessitore, "What makes people share political content on social media? The role of emotion, authority and ideology," *Comput. Human Behav.*, vol. 129, no. December 2021, p. 107150, 2022, doi:10.1016/j.chb.2021.107150.
- [45] K. Clark et al., "Advancing the ethical use of digital data in human research: challenges and strategies to promote ethical practice," *Ethics Inf. Technol.*, vol. 21, no. 1, pp. 59–73, 2019, doi:10.1007/s10676-018-9490-4.