





ORIGINAL

Detecting Public Issues in the 2024 East Java Gubernatorial Election through LDA Topic Modeling on Social Media X

Detección De Problemas Públicos En Las Elecciones A Gobernador De Java Oriental De 2024 Mediante El Modelado De Temas De LDA En Redes Sociales X

Aries Dwi Indriyanti¹ , Rahmat Gernowo² , Eko Sedyono³ , Mahrus Ali⁴ 

¹Diponegoro Univesity, Doctoral of Information System Department. Semarang, Indonesia.

²Diponegoro Univesity, Physics Departement. Semarang, Indonesia.

³Kristen Satya Wacana Univesity, Information Technology Department. Salatiga, Indonesia.

⁴Brawijaya Univesity, Information System Departement. Malang, Indonesia.

Cite as: Indriyanti AD, Gernowo R, Sedyono E, Ali M. Detecting Public Issues in the 2024 East Java Gubernatorial Election through LDA Topic Modeling on Social Media X. Data and Metadata.2025; 4:1140. <https://doi.org/10.56294/dm20251140>

Submitted: 10-11-2024

Revised: 31-03-2025

Acepted: 25-08-2025

Published: 26-08-2025

Editor: Dr. Adrián Alejandro Vitón Castillo 

Corresponding Author: Aries Dwi Indriyanti 

ABSTRACT

General elections are one of the pillars of democracy, so understanding public issues developing in the digital space is crucial to strengthening the legitimacy and quality of their implementation. The 2024 East Java gubernatorial election presents a strategic opportunity given the high level of regional political dynamics and the use of social media as a platform for political discourse. This study aims to identify key issues in public conversations on social media platforms related to the election, map their thematic distribution, and illustrate the distribution of public opinion across each topic. The research method used topic modeling analysis based on Latent Dirichlet Allocation. Data was collected from 3500 public posts relevant to the 2024 East Java gubernatorial election. The analysis process included text data cleaning, tokenization, removal of common words, and model development to obtain optimal topics. The results showed a model coherence value of 0,51, with three main topics: first, the implementation of regional head elections, including technical aspects, regional context, and the role of election organizers. Second, the security and smoothness of the election process, including voter management and trust in the results. Third, voter participation and the role of election organizers in ensuring legitimacy. The distribution of topics varies by region, influenced by local political backgrounds and previous election experiences. In conclusion, social media is a crucial arena for shaping public opinion and disseminating political issues. Digital data-based thematic analysis can help election organizers and policymakers design more effective public communications, increase participation, and strengthen public trust in elections at the regional level.

Keywords: 2024 East Java Gubernatorial Election; Public Opinion; Social Media; Topic Analysis; Thematic Distribution; Voter Participation; Election Legitimacy.

RESUMEN

Las elecciones generales son uno de los pilares de la democracia, por lo que comprender los problemas públicos que se desarrollan en el espacio digital es crucial para fortalecer la legitimidad y la calidad de su implementación. Las elecciones a gobernador de Java Oriental de 2024 representan una oportunidad estratégica dada la alta dinámica política regional y el uso de las redes sociales como plataforma para el discurso político. Este estudio busca identificar temas clave en las conversaciones públicas en plataformas de redes sociales relacionadas con las elecciones, mapear su distribución temática e ilustrar la distribución de la opinión pública en cada tema. El método de investigación utilizó análisis de modelos de temas basado

en la asignación latente de Dirichlet. Se recopilieron datos de 3500 publicaciones públicas relevantes para las elecciones a gobernador de Java Oriental de 2024. El proceso de análisis incluyó la limpieza de datos de texto, la tokenización, la eliminación de palabras comunes y el desarrollo de modelos para obtener temas óptimos. Los resultados mostraron un valor de coherencia del modelo de 0,51, con tres temas principales: primero, la implementación de las elecciones de jefes regionales, incluyendo aspectos técnicos, el contexto regional y el papel de los organizadores electorales. En segundo lugar, la seguridad y la fluidez del proceso electoral, incluyendo la gestión del electorado y la confianza en los resultados. En tercer lugar, la participación electoral y el papel de los organizadores electorales para garantizar la legitimidad. La distribución de los temas varía según la región, influenciada por los contextos políticos locales y las experiencias electorales previas. En conclusión, las redes sociales son un espacio crucial para moldear la opinión pública y difundir temas políticos. El análisis temático basado en datos digitales puede ayudar a los organizadores electorales y a los responsables políticos a diseñar comunicaciones públicas más eficaces, aumentar la participación y fortalecer la confianza pública en las elecciones a nivel regional.

Palabras clave: Gobierno del Estado 2024; Opinión Pública; Medios Sociales; Análisis Temático; Distribución Temática; Participación Ciudadana; Legitimidad Electoral.

INTRODUCTION

November 27, 2024, is a very important day for the Indonesian nation because the Regional Head Elections (Pilkada) will elect leaders in various regions, such as governors, regents, and mayors, at the same time,^(1,2) as well as being a concrete manifestation of participatory democracy.⁽³⁾ In addition, this process serves as a means to improve the quality of public services.^(4,5) and strengthen transparent and accountable regional governance.⁽⁶⁾ The Pilkada is also a test of the political maturity of the community, which is required to vote based on rational considerations rather than primordial sentiments.⁽⁷⁾

Gubernatorial elections have strategic significance because they determine the direction of policy and development at the provincial level. The 2024 East Java gubernatorial election was chosen as the focus of this research because the province has the second-largest population in Indonesia,⁽⁸⁾ high level of social, cultural, religious, and economic diversity,⁽⁹⁾ high level of social, cultural, religious, and economic diversity.⁽¹⁰⁾ This condition presents an opportunity to analyze political strategies, campaign patterns, public responses, and gender dynamics in the provincial-level contest.

Technological developments have transformed campaign strategies, from a predominance of face-to-face meetings to a digital approach through social media.⁽¹¹⁾ The East Java General Elections Commission (KPU) has also encouraged campaign innovation through activities such as the Peaceful Election Carnival.⁽¹²⁾ The history of social media use in political activities is also in line with the 2014 and 2019 Indonesian Presidential Elections, when Twitter, Facebook, and Instagram were intensively utilized to shape candidate images, disseminate political messages, and attract the support of young voters.⁽¹³⁾ However, their effectiveness remains debated. Not all candidates are able to build closeness with voters in the digital space.⁽¹¹⁾

Digitalization makes it possible to obtain information more quickly and widely and to use flexible tactics to boost voting, particularly among younger voters.⁽¹⁴⁾ Research in Semarang City conducted by Rizky showed that digital campaigns had a significant impact on voter participation aged 19-25 years, contributing 17,8 % to the variation in participation.⁽¹⁵⁾ A study in Bogor City by Rizatul et al.⁽¹⁶⁾ found that social media also shapes patterns of community involvement in digital democracy. Meanwhile, research on young cadres of the Perindo Party by Nurussyahadah, S et al.⁽¹⁷⁾ indicated that social media effectively enhanced public interaction and raised awareness of the party's political agenda.

International research also supports these findings. Younus et al.⁽¹⁸⁾ in a case study of Pakistan, revealed that digital-based campaigns through content creation, social media, influencers, and data targeting proved more effective than traditional methods. Based on these findings, the study chose platform X as the object of study due to its broad user base, high interaction reach, and effectiveness in shaping public opinion. Platform X, formerly known as Twitter, is a microblogging-based social media launched in 2006. In Indonesia, X is ranked 14th globally and 4th nationally with 27,1 million active users in early 2025.⁽¹⁹⁾

The effectiveness of X as a data source is supported by features such as trending topics, retweets, and replies that enable rapid information dissemination and real-time public discussion. Moreover, previous research demonstrates this not only within the political sphere but also across various other domains. In the political realm, research by Vigna-Gomez et al. utilized 15 million tweets to predict the results of legislative elections in Mexico,⁽²⁰⁾ Becatti et al.⁽²¹⁾ used 10 million tweets to characterize the behavior of automated accounts on Twitter during the 2019 UK elections, and Ansari et al.⁽²²⁾ analyzed Twitter user sentiment towards the main national political parties participating in the election process.

In the health sector, 1,45 million tweets successfully identified 62 000 breast cancer patients to map drug side effects.⁽²³⁾ In the disaster sector, this study analyzed 8900 interactions on X after the 5,1 magnitude earthquake in Oklahoma on February 2, 2024, covering seven days after the event, to demonstrate the important role of digital platforms in disaster response, while other research shows that X data is effective for post-earthquake situational awareness,⁽²⁴⁾ including analyzing public interactions, mapping affected locations, and identifying issues in real time, provided that it is processed with the right analysis methods.

Twitter provides a very large volume of data with informal, unlabeled, and highly dynamic characteristics, making manual or supervised learning approaches difficult to implement efficiently and highly dependent on large human resources. In this context, Latent Dirichlet Allocation (LDA), introduced by Blei et al. in 2003, presents a solution through a probabilistic topic modeling approach that views each document as a mixture of topics and each topic as a probability distribution of words.⁽²⁵⁾ By using a probabilistic topic modeling approach,⁽²⁶⁾ LDA effectively identifies underlying topics in unstructured text without the need for manual labeling.

LDA can also be used in conjunction with measures like coherence score to maximize the quantity of semantically significant subjects, particularly with social media data that is dynamic and diverse. LDA consistently produces better results (LDA coherence score of 0,580 compared to NMF coherence score of 0,500) in COVID-19-related study, which is consistent with studies by Mifrah S et al. Coherence score is a crucial parameter for evaluating topic models.⁽²⁷⁾ To extract hidden topics from huge corpus volumes, further studies have demonstrated that LDA performs better than LSA and PLSA when hyperparameter adjustment is used.⁽²⁸⁾

The urgency of this research lies in the need for a comprehensive mapping of public issues ahead of the 2024 East Java gubernatorial election, considering that social media X has become the main arena for the battle of political narratives. Three primary concerns are addressed by this study: (1) what were the primary subjects of public discourse on X in relation to the 2024 East Java governor election; (2) how the issues were distributed thematically; and (3) a map of public opinion on each topic.

It is innovative because it visualizes the geographic distribution of issues and sentiment by combining spatial analysis with the Latent Dirichlet Allocation (LDA) method. Identifying and categorizing public concerns, creating sentiment maps, and providing data-driven suggestions to campaign managers and legislators are among the goals of the study. In order to identify and track the dynamics of political debates in the digital sphere throughout the election process, this study focuses on LDA topic modeling on social media X.

METHOD

Types of Research

This research is a descriptive observational study with a quantitative-qualitative approach. The quantitative approach was used to model topics using the Latent Dirichlet Allocation (LDA) algorithm and analyze the frequency of keyword occurrences, while the qualitative approach was used to interpret the meaning of each resulting topic. This study did not intervene with the research object, but rather observed data available on social media platform X (Twitter).

Population and Sample

The research universe is all tweets published on social media X (Twitter) containing the keywords “PILKADAJATIM2024” and “PEMILUJATIM2024” in the specified time period. The research sample was taken using a purposive sampling technique based on certain search parameters, including locations in the East Java Province area with a search radius of 20-30 km from all cities and districts, Indonesian language, type of tweets in the form of original tweets (excluding retweets and replies), and the time span of data collection according to research needs.

Research Variables

The main variable of this study is the topic of public discussion regarding the 2024 East Java gubernatorial election on social media X, while the supporting variables include the frequency of keyword appearance and location distribution as supporting data for topic interpretation.

Data Collection

Data collection was conducted through web crawling techniques using the Twitter API with the assistance of a third-party library, TweetHarvest. The search parameters used refer to keywords, language, tweet type, time range, and location constraints, as described in the universe and sample sections.⁽²⁰⁾ The crawled data is automatically saved in CSV format for further analysis.

Research Procedure

This research procedure was carried out to achieve the objectives. The process consists of three main stages: data preprocessing, topic modeling, and data analysis.

Data Preprocessing

Most of Twitter(X) data is unstructured. Therefore, the initial data that is diverse or unstructured is processed into structured data so that it can be analyzed using the text mining method. Tweets obtained through the crawling process contain various obstacles, such as typos, use of slang, and grammatical errors. Therefore, a data pre-processing process is needed to ensure data quality before further analysis. In the initial stage, the data taken is automatically stored in a database in CSV format. After that, a series of pre-processing stages are carried out, including:⁽²¹⁾ Data cleansing is cleaning up irrelevant tweet data so that it becomes relevant data or contains unnecessary elements, such as URLs, mentions (@username), hashtags, numbers, symbols, and other special characters, so that the data becomes cleaner and more relevant for analysis. Tokenization is dividing sentences into sections called tokens. Tokens can be formed in words, phrases or other meaningful elements.

⁽²²⁾ Stopwords removal Removing common words that often appear but do not contribute significantly to the meaning of the sentence, such as “dan”, “yang”, “di”, and so on. The list of stopwords used is adjusted to Indonesian. Stemming, which is the process of getting the word base by eliminating affixes and suffixes. Slang Word Removal (Non-Standard Word Removal) to match the words in the text with the Big Indonesian Dictionary (KBBI) and correct or replace non-standard words into standard words. For example, the word “gk” or “ga” is changed to “tidak”.

Topic Modeling

Topic modeling: at this stage, the processed data is used in the topic modeling process using the Latent Dirichlet Allocation (LDA) algorithm. The LDA topic model is an unsupervised machine learning.⁽²³⁾ This model is useful in identifying hidden information in large document collections.⁽²⁴⁾ The topic modeling process begins with the transformation of preprocessed text data into a numeric representation using the Bag-of-Words (BoW) approach through the gensim.corpora module in Python. This representation produces a term-document matrix which is the main input for the Latent Dirichlet Allocation (LDA) algorithm. LDA works on the assumption that each document is a mixture of several topics, and each topic is a probability distribution over a set of words.^(25,26) The technical implementation uses the gensim.models library using the LdaModel package with parameter optimization through: (1) determining the number of topics (k) based on the coherence score (20) tuning the hyperparameters α (document-topic distribution) and β (topic-word distribution) automatically following the recommendations of Zhao et al.⁽²⁷⁾ for social media analysis. The model training process involves Gibbs sampling iterations to maximize the likelihood of topic distribution. The modeling results in the form of topic-document and word-topic probability matrices are then ready for the validation stage and interactive visualization using pyLDAvis.

Visualitation and Data Analysis

Bag of Words (BoW) is used to model text documents by calculating the frequency of occurrence of each word, ignoring word order and sentence structure.^(28,29) This representation produces a term-document matrix which is the main input in topic modeling using Latent Dirichlet Allocation (LDA).⁽³⁰⁾ After text processing, the results of LDA modeling are analyzed and visualized interactively using PyLDAvis, a Python library that allows exploration of word distribution and its relevance to each topic.⁽³¹⁾ To enrich the interpretation, various visualization techniques are used such as word clouds (displaying keyword frequencies) and bar charts (comparing topic dominance).

RESULTS AND DISCUSSION

This section describes the results of the topic analysis based on the method described in Section 2. First, a detailed description of the dataset containing public tweets related to the 2024 East Java gubernatorial election will be presented. Next, the results of this analysis are explored to reveal the main issues in public discussions and priority topics that dominate conversations on social media X. The classification and topic modeling process was carried out using minimal hardware in the form of a computer or laptop with an Intel Core i5 processor, 8 GB of memory, and 500 GB of SSD storage, and utilizing Python and Visual Studio Code software for code development and organization.

Data Collection Results

Data was collected from the social media platform X (formerly Twitter) on January 20, 2025, using a crawling method supported by the Twitter API and third-party libraries such as TweetHarvest. The dataset consisted of Indonesian-language tweets originating from cities and regencies in East Java Province, filtered using the keywords “PEMILU”, “PEMILIHAN GUBERNUR JAWA TIMUR”, and “PILKADA JATIM 2025”, resulting in 3771 tweets. After the data cleaning process, a total of 3600 tweets were ready for analysis, as summarized in table 1.

topics. The experimental results consistently showed that the model with two topics provided the best balance between coherence and information detail. So it can be concluded that choosing 2 topics will produce a model with the best quality in terms of coherence.

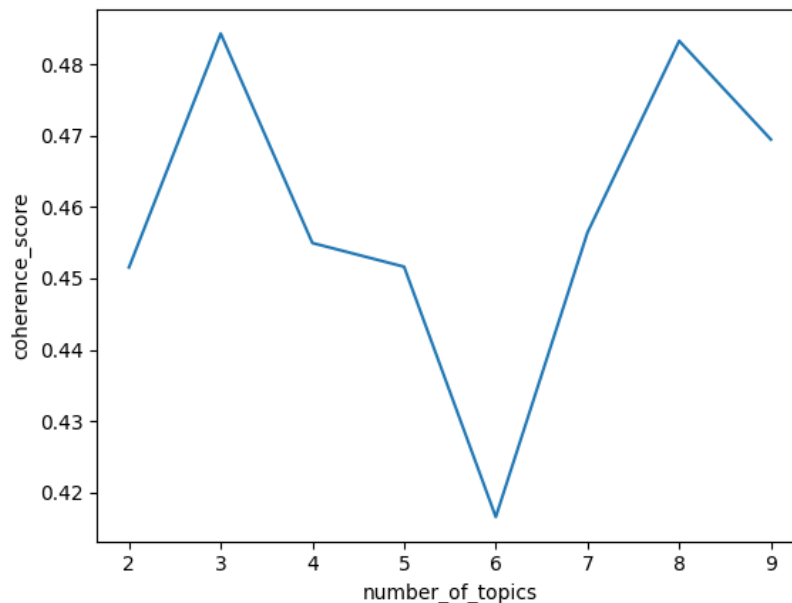


Figure 2. The coherence score of different number topics

Using the Gensim module to build a bag-of-words model from pre-processed data. Each document in the corpus is represented as a word vector, where each vector dimension represents the frequency of occurrence of a word in the document. This process begins with the formation of a dictionary that maps each unique word to a numeric index.

The LDA model in this study involves determining the Number of topics as much as 2 and the word of topic as much as 10. The modeling results produce a probabilistic distribution of words on each topic, with the highest probability words as topic representations. Table 2 presents the top keywords for each topic. The selection of these parameters, namely the number of topics and the length of words per topic, has significant implications for the quality and interpretability of the modeling results, so it requires careful consideration.

The parameters used as a reference to produce the best topic model are number of topics and words in topic. Probability is the number of times a word appears in a document. Word selection is taken as many as the number of words (word of topic) that have the highest probability. The following table 2, the three main topics generated are:

Topic	Probability * Word
1	0,059* <i>jawa</i> " + 0,049* <i>timur</i> " + 0,045* <i>daerah</i> " + 0,044* <i>kepala</i> " + 0,043* <i>pemilihan</i> " + 0,023* <i>kpu</i> " + 0,021* <i>serentak</i> " + 0,021* <i>pilih</i> " + 0,021* <i>hasil</i> " + 0,019* <i>selenggara</i> "
2	0,060* <i>daerah</i> " + 0,056* <i>kepala</i> " + 0,045* <i>pemilihan</i> " + 0,037* <i>jawa</i> " + 0,035* <i>timur</i> " + 0,017* <i>pilih</i> " + 0,013* <i>serentak</i> " + 0,012* <i>aman</i> " + 0,011* <i>laksana</i> " + 0,011* <i>suara</i> "
3	0,047* <i>timur</i> " + 0,046* <i>jawa</i> " + 0,044* <i>daerah</i> " + 0,044* <i>kepala</i> " + 0,042* <i>pemilihan</i> " + 0,041* <i>kpu</i> " + 0,025* <i>jatim</i> " + 0,023* <i>pilih</i> " + 0,021* <i>tahun</i> " + 0,021* <i>serentak</i> "

The issue of regional head elections in East Java is central to the three topics that were taken from the dataset, according to the topic modeling results.

Topic 1: implementation of Regional Head Elections in East Java

Topic 2: security Issues and the Election Implementation Process

Topic 3: KPU Involvement and Voter Participation

Topic 1 highlights the implementation of the elections, as seen from keywords such as "*Jawa*", "*Timur*", "*daerah*", "*kepala*", "*pemilihan*", "*KPU*", and "*selenggara*". This indicates that the discussion focuses on the regional context (East Java) and the organization of the local elections, including references to the election commission (KPU) and election outcomes.

Topic 2 also centers around the regional head elections but places more emphasis on the security and execution process, as indicated by words like “*aman*” (safe), “*laksana*” (implementation), and “*suara*” (vote). This suggests that the conversation in this topic touches on how the elections were conducted in a secure and systematic manner, ensuring proper vote management.

Topic 3 maintains topical relation to the preceding two and focuses on the institutional role of the KPU and the concurrent nature of the elections. Keywords such as “*KPU*”, “*Jatim*”, “*tahun*”, and “*serentak*” suggest a focus on the role of the election commission and the annual or scheduled nature of the regional elections, especially those held concurrently across regions.

Visualization and Data Analysis

Figure 3 shows the results of the visualization of topic modeling results using the pyLDAvis library displays the distribution and relationships between topics as circles in the left panel. The circle size indicates the topic’s dominance in the dataset. The right panel displays the 30 most common salient terms that differentiate one topic from another. Words such as Java, East, Region, Head, Election, KPU, and Serentak indicate that the dominant topics are related to the regional elections in East Java.

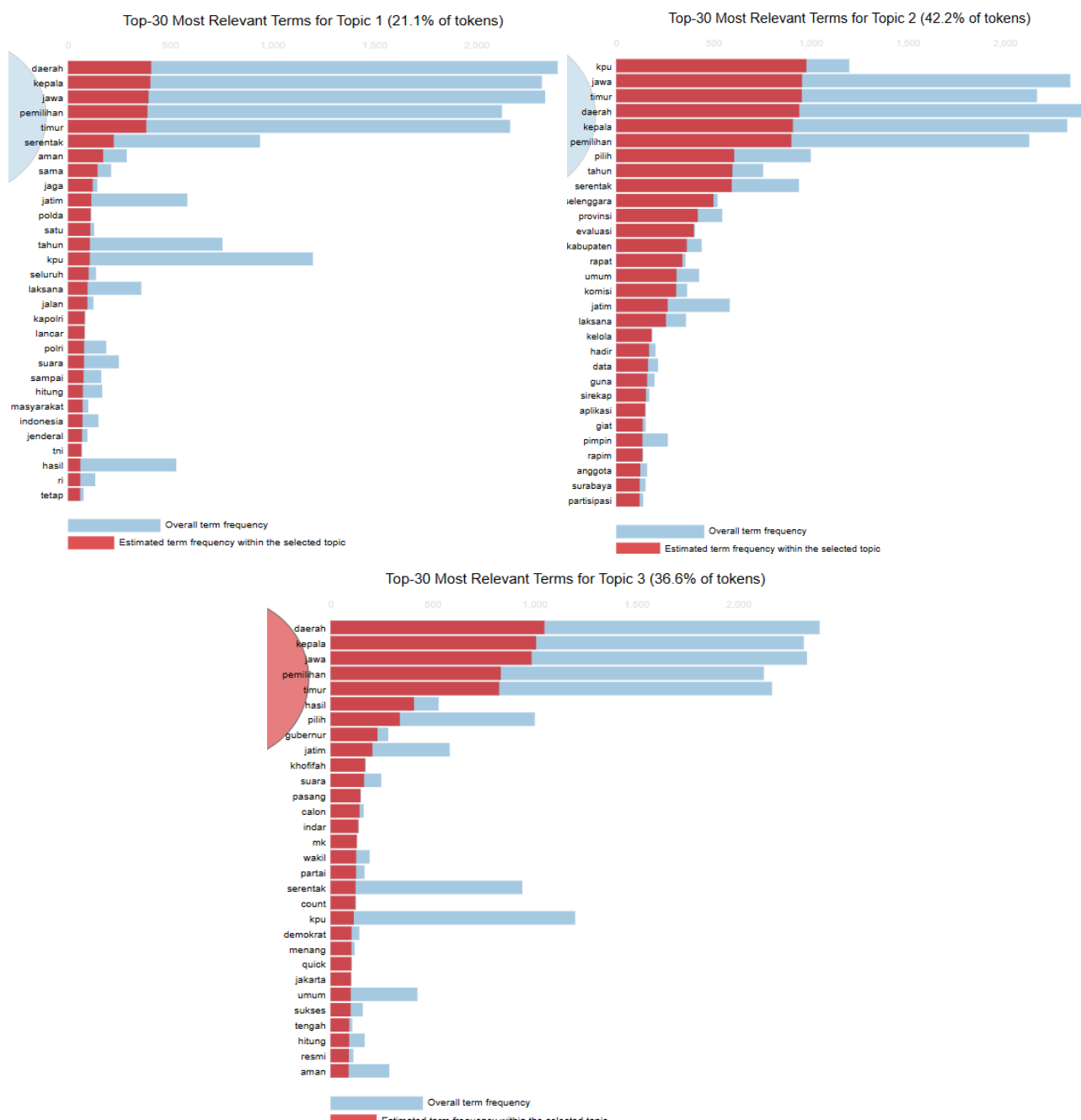


Figure 3. Term clustering visualization (Topics 1, 2, 3)

The interaction between the two panels helps understand the relationships between topics and the keywords that form them. Furthermore, five preprocessed documents or tweets were also clustered based on the LDA model, and the results of this clustering are presented in table 3:

Table 3. Document or Tweet classification	
Tweets	Topic
Bukan pemilihnya yg oon. Tetapi suara pemilih yg tidak memilih Khofifah di manipulasi. Pilkada Jatim .Jateng. dan Sumatera Utara persis cara caranya seperti pilpres 2024. Di Jawa timur ada di beberapa wilayah kabupaten suara Bu Risma Nol. Keanehan pilkada sama dg pilpres.	1
Ini Pilkada tahun 2015 hampir sepuluh tahun yang lalu di Jawa Timur. 2024 gimana? Kalau sekilas Pilkada 2024 ini di berbagai daerah banyak petahana yang tumbang. Rakyat pingin perubahan? Mau cari data mau nulis tapi duh duh dudududu https://t.co/4Rvf11yeqU	2
Komisi Pemilihan Umum (KPU) Jawa Timur telah menyelesaikan rekapitulasi perolehan suara Pilkada Jawa Timur 2024. Hasilnya menunjukkan pasangan calon Gubernur dan Wakil Gubernur Khofifah Indar Parawansa - Emil Elestianto Dardak unggul. https://t.co/ns6OyEQkD6 https://t.co/Ahq3VyvTpH	3

The research results show that public discourse on social media platform X leading up to the 2025 East Java regional elections was heavily influenced by issues related to regional head elections. The frequency of terms such as “KPU”, “pemilihan”, and “serentak” confirms that the public places significant attention on the election mechanism and its organizing institutions.

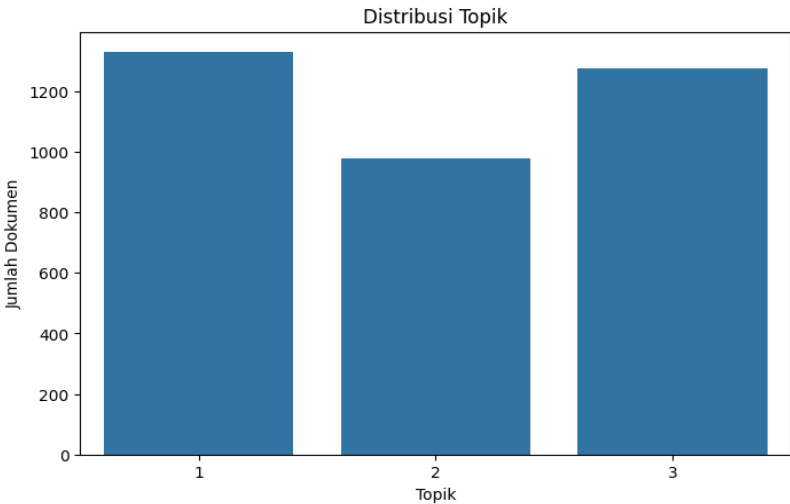


Figure 4. Visualize the Spread of Each Topic on Tweets Data

In figure 3 the following is a visualization or overview of the distribution of data from each topic. The y-axis or ordinate is the number of topic spreads, and the x-axis or abscissa is the number of topics corresponding to various topics. Shows the distribution of tweets for each topic. Topic 1 dominated with 1330 tweets, followed by topic 3 (1274 tweets) and topic 2 (977 tweets).

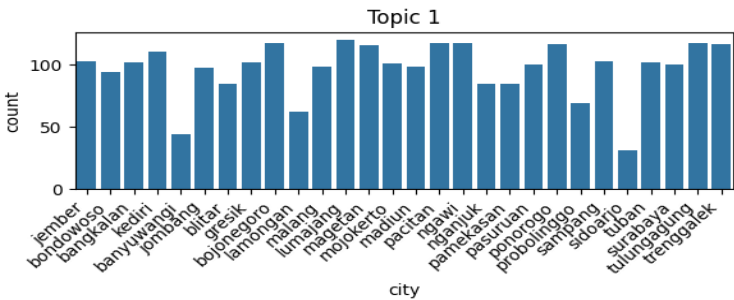


Figure 5. Visualization of the Distribution of topic 1

Figure 5 displays the results show that the majority of tweets discussed the general The visualization in figure 5 shows that topic 1, which discusses the general implementation of the regional elections in East Java and the role of the General Elections Commission (KPU) as the organizer, was the most frequently discussed topic. Cities such as Malang, Kediri, Tulungagung, and Blitar recorded a high number of tweets related to this topic, reflecting high public awareness of the simultaneous regional elections in their regions. Meanwhile, cities like Pacitan demonstrated lower levels of participation in this discussion.

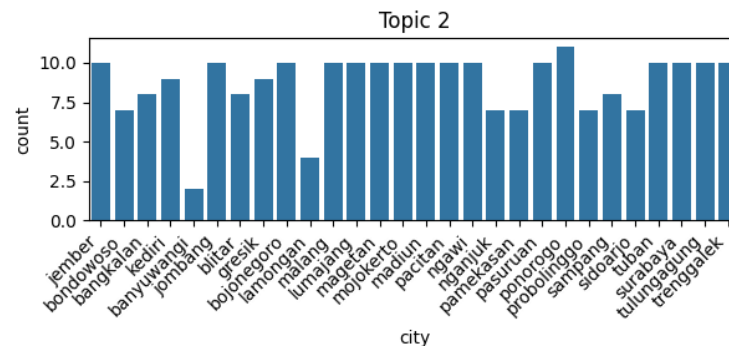


Figure 6. Visualization of the Distribution of topic 2

Meanwhile, figure 6 shows the distribution of topic 2, which relates to security issues and the technical processes of the regional elections. This topic was fairly evenly distributed across cities, with Tulungagung, Kediri, Malang, and Ngawi exhibiting a modest increase in discussion frequency. Conversely, Blitar and Pacitan appeared to have low engagement on this issue.

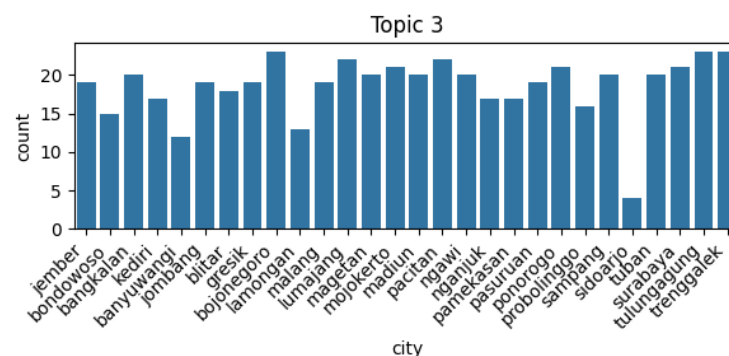


Figure 7. Visualization of the Distribution of topic 3

In figure 7, the distribution of topic 3, which highlights KPU involvement and voter participation, shows more significant variation across regions. Trenggalek, Tulungagung, and Sidoarjo were noted as the regions with the highest frequency of discussion on this topic. In contrast, Surabaya showed the lowest level of discussion, indicating minimal public attention to the aspect of election participation in the city.

DISCUSSION

Latent Dirichlet Allocation (LDA) analysis shows that public discourse is divided into three main issue groups. First, issues concerning the implementation of regional head elections, encompassing technical discussions of the implementation, the regional context of East Java, and the role of the General Elections Commission (KPU) as the organizer.⁽³²⁾ Second, issues focusing on the security and smoothness of the election process, including voter management and trust in the election results.⁽³³⁾ Third, issues related to the KPU's institutional involvement and voter turnout, reflecting public concern about the legitimacy and effectiveness of the simultaneous election administration.

The thematic distribution of these three issues shows variations in focus across regions. Election implementation issues were more prominent in areas with intensive campaign activity, while security issues received attention in areas with a history of disputes or discrepancies in vote results.⁽³⁴⁾ Meanwhile, topics concerning voter participation and the role of the KPU tended to be stronger in areas with low turnout or in areas serving as administrative centers. This variation demonstrates that local political backgrounds and previous election experiences influence the formation of public opinion on social media.⁽³⁵⁾

Mapping public opinion by topic reveals that perceptions of the elections in East Java are formed not only

by direct experience but also by interactions and discourses spread on social media. These findings align with previous research, which emphasized the crucial role of social media in shaping the public agenda and influencing public perceptions of the political process.⁽³⁶⁾ However, this study adds a new perspective by identifying thematic distribution and geographic variation of opinion, providing a more detailed picture of how election issues evolve in the digital space.

From the authors' perspective, these results demonstrate that social media-based topic analysis can be a strategic tool for election organizers and policymakers. Understanding dominant topics, their distribution, and the public opinion map can help design more effective public communication interventions, increase voter participation, and strengthen trust in the election process at the regional level.

CONCLUSIONS

This study aims to: (1) identify the main topics of public discourse on social media platform X regarding the 2024 East Java gubernatorial election, (2) analyze the thematic distribution of discussed issues, and (3) map public opinion on each topic. The method used was topic modeling with the Latent Dirichlet Allocation (LDA) algorithm. The research data consisted of public posts on social media platform X relevant to the 2024 East Java gubernatorial election. These data were then processed through text preprocessing, LDA model development, and topic coherence evaluation.

The results showed that the LDA model with three topics yielded the highest coherence value of 0,410638, indicating that the resulting topics had a fairly good sense of interconnectedness and could be clearly interpreted. The three main topics identified were: (1) Implementation of Regional Head Elections in East Java, (2) Security Issues and the Election Implementation Process, (3) KPU Involvement and Voter Participation. The first topic was the most dominant, followed by the second and third topics, whose distribution varied across regions.

Based on the author's analysis, differences in topic distribution indicate that the issues discussed by the public are influenced by local factors such as regional political dynamics and previous election experiences. This finding confirms the role of social media as a dynamic space for voicing political aspirations and as an indicator of public perception of the democratic process. Practically, these results can be utilized by the General Elections Commission (KPU), local governments, and policymakers to improve the effectiveness of political communication, expand voter participation, and strengthen the legitimacy of election administration.

However, this study has limitations because it only analyzed conversations on a single social media platform and focused on Indonesian, thus failing to account for potential variations in opinion across regional languages or other media. Future research is recommended to expand data sources, use a multimodal approach, and integrate sentiment analysis to obtain a more comprehensive picture of public opinion on the election.

BIBLIOGRAPHIC REFERENCES

1. Tahapan dan Jadwal Pemilihan Gubernur dan Wakil Gubernur, Bupati dan Wakil Bupati serta Walikota dan Wakil Walikota Tahun 2024. Komisi Pemilihan Umum (KPU). 2024. <https://infopemilu.kpu.go.id/Pemilihan/Pilkada>
2. Amir UA, Widyasamratri H. Peran Masyarakat dalam Pembangunan Wilayah. Sustain Plan Cult SPACE J Perenc Wil Dan Kota. 2022 Mar 30 [cited 2025 July 11];3(2):30-4. <https://ejournal.unhi.ac.id/index.php/space/article/view/2242>
3. Christina L, Kristanto SB. Mendeteksi Perilaku Fraud Laporan Keuangan Dengan Theory of Planned Behavior (Studi Empiris Pada Industri Perbankan). J Akunt Kontemporer. 2019;11(1):1-13.
4. Kim SY. Direction of Revision of Relevant Laws for Balanced Regional Development. Public Law J. 2022 Aug 31; 23(3):3-27. <http://www.dbpia.co.kr/Journal/ArticleDetail/NODE11125356>
5. Coonrod J. Participatory Local Democracy: Key to community-led rural development. Development. 2015 June;58(2-3):333-40. <http://link.springer.com/10.1057/s41301-016-0008-2>
6. Eberlei W. Democratic Governance in Sub-Saharan Africa. Ghana J Dev Stud. 2012 Feb 10 ;8(1). <http://www.ajol.info/index.php/gjds/article/view/73605>
7. Why It's Rational to Vote. In: Rationality, Democracy, and Justice. 1st ed. Cambridge University Press; 2015. p. 21-49. https://www.cambridge.org/core/product/identifier/9781107588165%23c06523-120/type/book_part
8. Jumlah Penduduk Menurut Provinsi di Indonesia (Ribu Jiwa), 2024. 2024. <https://sulut.bps.go.id/id/>

statistics-table/2/OTU4Izl=/jumlah-penduduk-menurut-provinsi-di-indonesia.html

9. Jatim Bisa Jadi Barometer Nasional dalam Hal Keberagaman. Dinas Kominfo Provinsi Jawa Timur. 2022 Desember 16:45:38; <https://kominfo.jatimprov.go.id/berita/jatim-bisa-jadi-barometer-nasional-dalam-hal-keberagaman>

10. Jatim KPU. Daftar Calon Tetap Pemilihan Umum (DCT Pemilu) Tahun 2024. KPU Jatim. 2024; <https://jatim.kpu.go.id/page/read/daftar-calon-tetap-pemilihan-umum-dct-pemilu-tahun-2024>

11. Prasetyo KB, Rahardjo T, Sardini NH. Digital Campaign Dysfunction in Indonesian Re-gional Election During the COVID-19 Pandemic. *Int J Res Rev.* 2024 Mar 30. 2025;11(3):345-59. https://www.ijrrjournal.com/IJRR_Vol.11_Issue.3_March2024/IJRR44.pdf

12. Rifansyah MR, Anisykurlillah R. Analisis Strategi Komisi Pemilihan Umum (KPU) Jawa Timur untuk Meningkatkan Partisipasi Politik Masyarakat dalam Menghadapi Pemilu Tahun 2024. *J Prog Adm Publik.* 2024;4(1):32-9. <https://jurnal.utb.ac.id/index.php/jpap/article/view/1218>

13. Sellita. Media Sosial dan Pemilu: Studi Kasus Pemilihan Presiden Indonesia. *J Lemhan-nas RI.* 2022;10(3):1-16. <http://jurnal.lemhannas.go.id/index.php/jkl/article/view/293>

14. Taras D, Davis R. Electoral Campaigns, Media, and the New World of Digital Politics. Ann Arbor, MI: University of Michigan Press; 2022. <https://www.fulcrum.org/concern/monographs/gx41mk97m>

15. Rizky M. Efektivitas Kampanye Digital Dalam Meningkatkan Partisipasi Pemilih Muda Pada Pemilihan Kepala Daerah Kota Semarang 2024. *J Audience.* 2025 Feb 1 [cited 2025 Aug 10];7(2):182-200. <https://publikasi.dinus.ac.id/index.php/audience/article/view/11865>

16. Rizanul RC, Egistin D, Jasmine A, Juniarto D. Analisis Peran Media Sosial (TikTok) dalam Dinamika Partisipasi Politik pada Pemilu Presiden 2024. *J Soc Contempl.* 2025 Feb 28 [cited 2025 Aug 10];3(1):51-61. <http://www.journal.idereach.com/index.php/JSC/article/view/91>

17. Nurussyahadah S, Rasyadi AH, Ayu NR, Herliany SD, Mafaza T. Strategi Kampanye Digi-tal : Studi Kasus Pemanfaatan Media Sosial Oleh Kader Muda Perindo. *J Ilmu Sos Dan Ilmu Polit.* 2025;6(2):97-114. <https://journal.stisipolrajahaji.ac.id/index.php/jisipol/article/view/193>

18. Younus M, Nurmandi A, Mutiarin D, Luhur Prianto A, Abdul Manaf H, Irawan B. Running Digital Political Marketing Movement for Election 2024: A Case Study of Pakistan. *J Polit Mark.* 2024;1-23. <https://www.tandfonline.com/doi/full/10.1080/15377857.2024.2438379>

19. Kamp S. X Users, Stats, Data & Trends for 2025. DataReportal. 2025. <https://datareportal.com/essential-x-stats>

20. Röder M, Both A, Hinneburg A. Exploring the Space of Topic Coherence Measures. In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining.* ACM; 2015. p. 399-408.

21. Fitri VA, Andreswari R, Hasibuan MA, Fitri VA, Andreswari R, Hasibuan MA. Sentiment Analysis of Social Media Twitter with Case of Anti- Sentiment Analysis of Social Media Twitter with Case of Anti- LGBT Campaign in Indonesia using Naïve Bayes , Decision Tree , LGBT Cam-paign in Indonesia. 2019;00.

22. Kurniati ID, Setiawan R, Rohmani A, Lahdji A, Tajally A, Ratnaningrum K, et al. Buku Ajar Text Mining. Dijaya R, editor. UMSIDA Press Anggota IKAPI No. 218/Anggota Luar Bia-sa/JTI/2019 Anggota APPTI No. 002 018 1 09 2017 Redaksi; 2015.

23. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimed Tools Appl.* 2019 June [cited 2025 July 11];78(11):15169-211. <http://link.springer.com/10.1007/s11042-018-6894-4>

24. Rashif F, Nirvana GIP, Noor MA, Rakhmawati NA. Implementasi LDA untuk Pengelompo-kan Topik Cuitan Akun Bot Twitter bertagar #Covid-19. *Cogito Smart J.* 2021;7(1):1-12.

25. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. 2003;3:993-1022.
26. Blei DM. Probabilistic Topic Models. 2012;55:77-84.
27. Zhao F, Zhu Y, Jin H, Yang LT. A personalized hashtag recommendation approach using LDA-based topic model in microblog environment. *Future Gener Comput Syst.* 2016;65:196-206.
28. Putri WTH, Hendrowati R. Penggalan Teks Dengan Model Bag Of Words Terhadap Data Twitter. *J Muara Sains Teknol Kedokt Dan Ilmu Kesehat.* 2018;2(1):129-38.
29. Alfanzar AI, Rozas IS. Topic Modelling Skripsi Menggunakan Metode Latent. *JSil J Sist Inf.* 2020;7(1):7-13.
30. Sriurai W. Improving Text Categorization By Using A Topic Model. *Adv Comput Int J.* 2011;2(6):21-7.
31. Sievert C, Shirley KE. LDavis: A method for visualizing and interpreting topics [Inter-net]. Stanford University; 2014 June p. 63-70. <https://nlp.stanford.edu/events/illvi2014/papers/sievert-illvi2014.pdf>
32. Qomariyah S, Iriawan N, Fithriasari K. Topic modeling Twitter data using Latent Dirichlet Allocation and Latent Semantic Analysis. In Surakarta, Indonesia; 2019. p. 020093. <https://pubs.aip.org/aip/acp/article/819632>
33. Placek M. Social media, quality of democracy, and citizen satisfaction with democracy in central and eastern Europe. *J Inf Technol Polit.* 2024 Jan 2;21(1):6-22. <https://www.tandfonline.com/doi/full/10.1080/19331681.2023.2220319>
34. Muhtar M, Setiyono B, Fitriyah F, Sardini NH. Political Polarization In The 2024 Local Elections: An Analysis Of Its Impact And Implications For Local Democracy In Indonesia. *JWP J Wacana Polit.* 2025 Jan 14;10(1):41-51. <https://jurnal.unpad.ac.id/wacanapolitik/article/view/58474>
35. Sellita. Media Sosial dan Pemilu: Studi Kasus Pemilihan Presiden Indonesia. *J Lemhan-nas RI.* 2022 Oct 9;10(3):1-16. <http://jurnal.lemhannas.go.id/index.php/jkl/article/view/293>
36. Steinveg B, Bjørnå H. Social media as an agenda-setting instrument in local politics. *Nord Rev.* 2023 June 1;44(2):299-314. <https://www.sciendo.com/article/10.2478/nor-2023-0016>

FINANCING

The authors did not receive financing for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Aries Dwi Indriyanti.
Data curation: Aries Dwi Indriyanti.
Formal analysis: Aries Dwi Indriyanti.
Research: Aries Dwi Indriyanti.
Methodology: Aries Dwi Indriyanti.
Project management: Aries Dwi Indriyanti.
Resources: Rahmat Gernowo.
Software: Aries Dwi Indriyanti.
Supervision: Rahmat Gernowo.
Validation: Eko Sedyono.
Display: Aries Dwi Indriyanti.
Drafting - original draft: Eko Sedyono.
Writing - proofreading and editing: Eko Sedyono.