

Topic modelling analysis of public policy narratives on prabowo-gibran in national news

Lukmanul Hakim¹, Anggi Yudistira Aditya², Muharman Lubis³

^{1,2}Faculty of Technic and Business, Universitas Madani, Indonesia

³Department of Information System, Universitas Telkom, Indonesia

Article Info

Article history:

Received December 10, 2025

Revised December 21, 2025

Accepted December 27, 2025

Keywords:

Topic modelling

Public policy

Prabowo-Gibran

Online news

Big data

ABSTRACT

The rapid acceleration of digital transition has become an inevitable reality of the modern era. The proliferation of online communication platforms, news portals, and heterogeneous data formats has substantially increased big data volumes, leading to large-scale collections of unstructured data. This study aims to analyze dominant public policy-related topics concerning the Prabowo-Gibran administration by applying topic modeling techniques to national online news media. Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) were employed as unsupervised learning approaches to extract latent semantic structure from a corpus of 200 credible news articles collected through URL fetching using Python 3. Data preprocessing included text cleaning, tokenization, bigram and trigram construction, and the development of a dictionary and corpus. Model performance was evaluated using topic coherence metrics, yielding scores of 0.3709 for LDA and 0.68 for NMF. To examine temporal dynamics, the dataset was divided based on the official inauguration date of the president and vice president, enabling a comparative analysis of dominant topics before and after the inauguration. Topic similarity across both periods was measured using cosine similarity, with the highest similarity score of 0.663 observed between Topic 4 in the pre-inauguration period and Topic 1 in the post-inauguration period. The findings provide insights into evolving media discourse and policy-related topic trends across the two periods, demonstrating the potentials of topic modeling in analyzing large-scale unstructured news data for diverse purposes to bridge computational science and empirical evidence of social science.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Lukmanul Hakim,

Faculty of Technic and Business,

Universitas Madani,

Email: lukmanul.hakim@umad.ac.id

<https://doi.org/10.52465/joscecx.v6i4.632>

1. INTRODUCTION

The evolving of digital platforms such as social media, online news portals, online communication forums, and others has rapidly triggered the availability of large, unstructured datasets which has potential to leverage big data analysis [1]. The public's demand for quick information and convenient access to news has driven to digital cumulative contents on a variety of topics, including other interesting data and facts that can

be utilized for various purposes, whether in the commercial, auditing, or academic sectors in an automatic or semi-automatic manners [2] specially to get feedbacks and public responses regarding government services in term of public policy arrangement improvement [1].

Actually, each online media platform has its own framing in presenting news, referring to Bagaskoro (2022); it can be seen how the media selects aspects of news based on reality in the field and highlights them in the text in the form of communication and promotion to the public [3]. As a result, the public receives accurate and credible information using language that is common, concise according to actual events, detailed explanations, including evaluations from the editorial perspective, and avoids the use of complicated vocabulary [4].

The role of media in the digital era has a significant impact on the public engagement. According to the Reuters Institute [5], 40% of people tend to avoid the news because it affects their mood and has an impact on readers' psychology regarding a news topic. In addition, 31% choose to avoid the news due to feeling overwhelmed and saturated with the amount of news continuously produced every day. This becomes a new challenge for readers, where there is a limitation in the perspective of information presented by the media against the time needed to read and the latest facts missed regarding an event [4]. However, overall media trust score in Indonesia remains stable and most people still considering news media as best sources for local politics or government across media markets [5].

In the context of Indonesia, the rollout of a new government typically generates extensive media attention particularly during February election [5], with diverse framings of both planned and forthcoming public policies throughout the governing period. Online news outlets continuously report on policy-related issues-particularly within symbolic timeframes such as the first 100 days of governance-by evaluating policy commitments in relation to the administration's stated visions and missions, from multiple perspectives shaped by journalistic norms. Consequently, it is not easy for people who have interest for news summary related policy [6].

With enormous amount of textual data from news outlets, there is a shift towards big data analytics rather than simple data retrieval by harnessing text mining to extract pattern from Natural Language Processing (NLP) in many sectors. The use of qualitative methods for text analysis, such as manual coding, discourse analysis methods, or grounded theory, still relies on labor-intensive manual procedures that seem to have reached their natural limits when dealing with large quantities of text [7]. The current challenge is how to extract relevant news sources from national media given the limited access due to the absence of an officially published Application Programming Interface (API) from national news outlets. It can be seen that most of Indonesian scholars still rely on web scraping refer to [6], [8] third-party news aggregator according to Saifudin (2021) [9], or manual according to study by [10] when dealing with national online news and Indonesian news texts unlike social media platforms which provide developer's tools for data crawling. Moreover, the process of extracting and analyzing web-based information from large, unstructured text datasets is a major challenge in the modern era, requiring the systematic use of information retrieval techniques to discover new meaningful features [11].

Some reviews on text mining applications have been conducted vastly in various contexts, such as the work by Hassani et al. (2020) [2], which performed extensive analyses within the scope of big data applied in both the industrial sector and academic studies. Meanwhile, Antons et al. (2020) [7] explored the extent of text mining's potential in innovative research across 124 journal articles from 10 journals in the field of innovation management and 8 top journals in general management. Wang & Lo (2021) [12] described resources to support text mining applications, especially during the COVID-19 pandemic where several hundred academic papers continue to be published every day that leads to difficulties of analysis and obtain latest findings. Topic models are a common approach used in text mining [7], [8]. Churchill & Singh (2021) [13] specifically discussed topic modeling, particularly unsupervised topic models, from their inception in the 1990s, the development of algorithms up to 2020, and their evolution over the years.

In addition, the application of topic modelling in policy-related studies was conducted by Rochmawati et al. (2022) [14] concerning the communication of COVID-19 health protocol policies on mask usage. In their paper, the authors collected Twitter data to be processed using topic modelling techniques and compared public perceptions before and after the issuance of presidential instructions. Raditya (2023) [15] utilized topic modelling techniques to identify strategic policy issues in the tourism investment sector by analyzing data from online news sources. Next, the study by Hidayatullah et al. (2021) [16] focused on exploring topics related to the National Strategic Projects (PSN) in national online media and employed a bigram model to facilitate interpretation and intuition during Jokowi's administration period. We get a couple similar studies talking about policy-related to Prabowo-Gibran [16]–[20] and using diverse data sources such as social media, Youtube comments and so on which is tend to explore public responses or discourses upon policies and public sentiments, not policy made by the government, viral events, popular figure like president, or very important person. We conclude that prior literatures, particularly in policy-related studies in Indonesia, most of them

combining news topic modelling to determine news sentiment upon policy, media framing analysis, and thematic analysis such as infrastructure, tourism, and policy-related covid 19.

Therefore, this study aims to analyze the public related-policies of the Prabowo-Gibran narratives in national online media. We collect data from various popular online news outlets such as detik.com, tempo.co, kompas.com, bbc.com, okezone.com and more news aggregations, which is then analyzed using topic modeling with different techniques which are LDA and NMF. We also comparing topics similarities and topic dominant before (pre) and after (post) ceremony as for narrative tracking, or shift towards policies across actors and time published on media. The analysis focuses on topics of public policy by leveraging data growth and techniques used in text mining to uncover hidden topics discussed in national news related to public policy. This is driven by the increasing demand for text mining and Natural Language Processing (NLP) to reveal hidden meanings from textual data in recent years [2], [21], and it has been applied in various contexts such as case studies, demonstrations, literature reviews, and variable exploration [7]. This study is valuable in media narratives about policy-related and giving opportunities to online news readers on topics changes over time.

2. METHOD

2.1. Data gathering and text preprocessing

A process of 200 data gathering carried out by searching for the keywords Prabowo-Gibran across various national news portals such as “Prabowo”, “Gibran”, and “Policy” with manual filters to obtain primary data sources. The target URLs then collected in order to be process with python library named “newspaper3k”. Data such as news titles, publication dates, URLs, and news contents were fetched and put it into a data frame, which was then cleaned using pre-processing methods like stemming with “Sastrawi” and using “nltk” to remove stopwords, normalize Indonesian text, and add custom stopwords in the program. The removal of non-letter characters was done using regular expressions, and the text format was converted into tokens and token n-grams such a study [21] by adding two new columns to the data frame. Most of pre-processing steps in this study referring to prior studies [8], [22], [23].

Next process is splitting data into two categories which are pre-inauguration and post-inauguration according to the official inauguration date of the Indonesian president and vice president on October 20, 2024. This data used for next process to be visualized with word clouds and compared using Cosine Similarities which is a simple and effective method [21] to measure the topic closeness for each period. Also used for policy trends observation by generating topic dominant graph to see the topic distribution in each period. We successfully categorized 125 documents for post-inauguration (pasca pelantikan), 61 for pre-inauguration (pasca pelantikan), and rest 14 undated. Both periods were visualized using word clouds by combining 10 topics in one word-cloud for pre-inauguration and post-inauguration respectively.

Table 1. Dataset examples and after cleaning

No	News title	Source	Raw data	Clean data
1	Janji Gibran Rakabuming Setelah Nanti Dilantik Jadi Wali Kota Solo	Detik.com	Gibran Rakabuming Raka-Teguh Prakosa resmi ditetapkan sebagai Wali Kota dan Wakil Wali Kota Solo terpilih. Gibran menjanjikan sejumlah program ketika sudah menjabat Wali Kota Solo.	gibran rakabuming raka teguh prakosa resmi tetap bagai wali kota dan wakil wali kota solo pilih gibran janji jumlah program ketika sudah jabat wali kota solo
2	Greenpeace Beri Prabowo-Gibran Nilai D di Bidang Transisi Energi	Tempo.co	TEMPO.CO, Jakarta - Greenpeace Indonesia memberi nilai D untuk pemerintahan Prabowo Subianto-Gibran Rakabuming Raka di bidang transisi energi. Greenpeace memberikan nilai tersebut untuk kinerja keduanya menjelang 100 hari pemerintahan.	tempo co jakarta greenpeace indonesia beri nilai d untuk perintah prabowo subianto gibran rakabuming raka di bidang transisi energi greenpeace beri nilai sebut untuk kerja dua jelang hari perintah
.....
200	Prabowo Tegaskan Pemerintah Serius Berantas Penyelundupan dan Tambang Ilegal	Kompas.com	JAKARTA, KOMPAS.com — Presiden RI Prabowo Subianto menegaskan pemerintah serius dalam memberantas kasus tambang ilegal. Hal tersebut disampaikan Prabowo saat	jakarta kompas com presiden ri prabowo subianto tegas perintah serius dalam berantas kasus tambang ilegal hal sebut sampai prabowo saat laku tinjau dan serah enam smelter

melakukan peninjauan dan penyerahan enam smelter timah sitaan Kejaksaan Agung ke PT Timah Tbk saat kunjungan kerja di Bangka Belitung

timah sita jaksa agung ke pt timah tbk saat kunjung kerja di bangka belitung

Before creating a dictionary and corpus, we need to form a contiguous sequence of n element in the dataset. The goal of this process is to enhance semantic meaning of phrases from the corpus as well as to remove tokens that occur too frequently and very rarely in the data frame, leaving inputs ready for training.

2.2 Topic modelling inference

Understanding the core subjects within a collection of documents is a very fundamental task in the information era [13]. LDA is one of generative probabilistic models [6] that has been widely used across various sectors, including the realm of public policy [24], in order to mine a number of documents to obtain the main topics that appear most frequently in a corpus, turning them into a summarized visualization in the form of topics or groups of words that are related each other [12], [22].

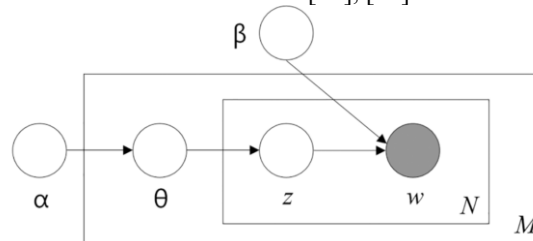


Figure 1. LDA model (Source: <https://doi.org/10.1371/journal.pone.0238972.g001>)

In Figure 1, the variable w or denoted as d represents the observed or selected words, namely the word or token, and z is the topics choices based on w , while the other variables (θ) are hidden (latent) variables estimated during the inference process. The box labeled M represents the number of documents, and N indicates the number of words per document [13], [21], [23].

In contrast, the application of Non-negative Matrix Factorization (NMF) in topic models has also been applied in various studies. This model is one of the non-probabilistic statistical models used to find factors or decompose a non-negative matrix into two non-negative submatrices [22], [23] and is conceptually formulated in figure 2.

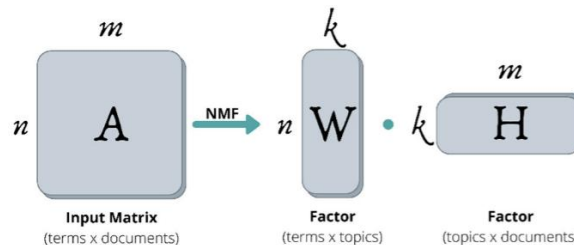


Figure 2. NMF model (Source: <https://doi.org/10.3389/fsoc.2022.886498>)

A represent the Term-Document Matrix where n is terms and m is documents, while W represent the number of tokens or words in the matrix columns. Each document is then made into a document and topic matrix to see how dominant each topic is within each document. Then, the matrix of each topic is searched with each word or token.

These two algorithms fall into the category of unsupervised machine learning and good general purpose topic models [21], which processes a set of documents D as input data and presents topic outputs T as a representation of the main discussion points of documents D in an accurate and coherent manner [13]. However, we have to labelled each topic based on author intuitive approach or judgement according to [6], [13], [22] including NMF [23]. This is expected in unsupervised learning method like LDA without pre-defined categories [7].

The input process for this study, performed by training the LDA model to instruct the model to read all documents 20 times with a 'passes' value of 20 and perform 'iterations' 400 times, aimed at finding the hidden topic likelihood in the corpus. This step including topic extraction and tuning hyperparameter to get word probabilities in a topic and topic probabilities in a document [25].

2.3. Topic evaluation

Topic quality can be determined using intuitive reasoning by looking at topic coherence, ease of interpretation, and having few overlapping words [13], [26]. In order to find the number of topics optimization, we conduct a coherence process by retuning hyperparameter of LDA.

2.4. Topic visualization and interpretation

We start the interpretation by visualizing histogram from text preprocessing step. This pace use to create a sequence of bigram and trigram model were applied to the clean data to obtain sequences of two or three words that have a strong relationship using the “phrases” function from gensim. This is aimed at finding word combinations that frequently appear at least 5 times throughout the entire dataset with a threshold of 10.

3. RESULTS AND DISCUSSIONS

It can be seen from the histogram; several compound words obtained within a number of documents as shown in figure 3. Some consecutive words dominantly related to Name Identity Recognition (NER) like a figure of person like Prabowo, Gibran, Joko Widodo, Try Sutrisno and their position in the government such as president, vice president, major, constitutional court, and so on. Another pair or triplet of consecutive words refer to an event such as military anniversary, election process, candidates, and related to a place (A city Solo).

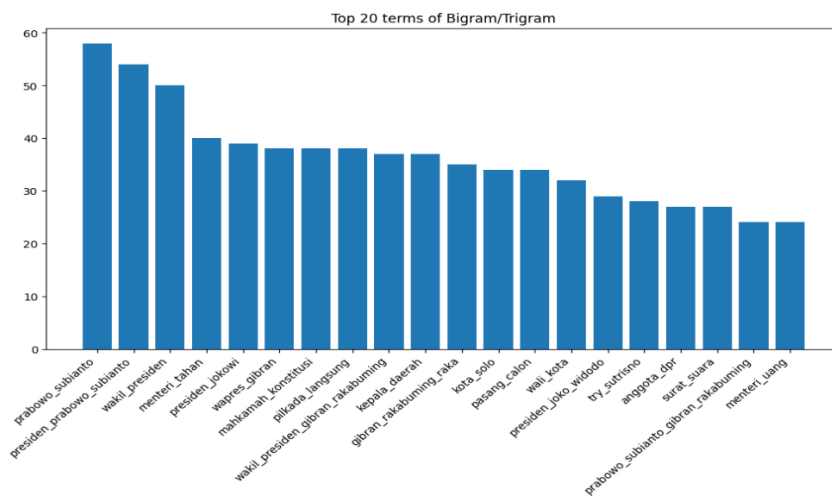


Figure 3. Consecutive words of n-gram

Table 2 showing about topics based on 10 keywords with the highest likelihood scores determining topic’s relevance with policy-related, indicating that higher scores imply more coherence and interpretable topics. We defined the topic number starting from topic 5 to 20. Based on our intuitive decision, we chose 5 topics that related to Gibran policy-related such as topic 3 but not clearly stated about specific program was planned or done during his period as Solo City Major. Another topic cluster 11 and 14 correlated with Prabowo-Gibran and their correspond with military, equipment, and defense industry. Meanwhile, topic 11 and 12 related to island and Indonesia new capital city.

Table 2. LDA topic cluster

Keywords based on topics	Topic label
Topic 3: 0.057*"gibran" + 0.031*"bangun" + 0.023*"kota_solo" + 0.019*"solo" + 0.016*"program" + 0.011*"persen" + 0.011*"wali_kota_solo" + 0.010*"masyarakat" + 0.010*"proyek" + 0.009*"kerja"	Gibran program, project for Solo
Topic 9: 0.020*"gibran" + 0.018*"negara" + 0.015*"milik" + 0.010*"kait" + 0.009*"tni" + 0.009*"lihat" + 0.009*"pres" + 0.009*"menteri" + 0.008*"wapres" + 0.008*"usul"	Gibran and Military
Topic 11: 0.031*"bijak" + 0.014*"masyarakat" + 0.013*"negara" + 0.012*"program" + 0.012*"publik" + 0.010*"kerja" + 0.008*"jalan" + 0.008*"putus" + 0.007*"hadap" + 0.007*"pulau"	Topic related to island and national program
Topic 12: 0.105*"ikn" + 0.044*"bangun" + 0.040*"wapres" + 0.037*"rencana" + 0.031*"selesai" + 0.029*"wapres_gibran" + 0.027*"gibran" + 0.025*"kantor" + 0.025*"kota_nusantara_ikn" + 0.022*"usul"	Gibran new office at Indonesia new capital city

Topic 14: 0.051*"tahan" + 0.035*"menteri tahan" + 0.020*"negara" + 0.018*"industri tahan" + 0.016*"menhan" + 0.015*"alutsista" + 0.013*"kuat_tahan" + 0.012*"angkat" + 0.012*"tni" + 0.011*"menteri tahan prabowo subianto" Prabowo, military equipment, and defense industry

Based on our evaluation, LDA model yielded a coherence score of 0.3709 with the top 5 topics, while the NMF model had a coherence score of 0.68. The comparison between the two shows a fairly significant score difference, indicating that iterative refinement is needed for LDA to produce more coherent topics.

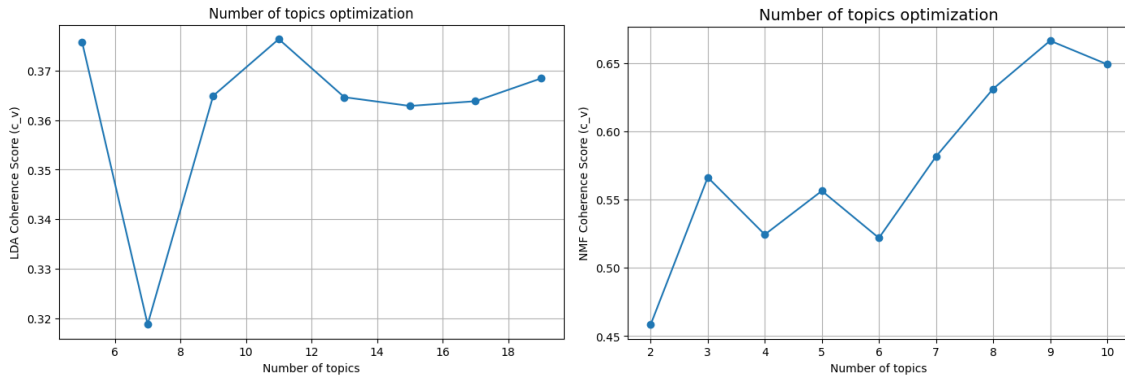


Figure 4. Result of topic coherence between models

The following word's structures show topics cluster performed by NMF. We chose the most relevant topics according to its matrix calculation and here it can be concluded that NMF more interpretable rather than LDA topic cluster. Moreover, two word-cloud comparisons were combined between pre and post inauguration. It can be seen from the pre-period dominated by word constitutional court, a city of Solo, commissioner, revitalization, and military equipment. Whereas, post inauguration word-cloud refer to IKN (Indonesia New Capital), and Name Entity Recognition frequently. Only few words correlated to economy, traditional market, law, corruption, and public housing.

- Topic 1: ['politik', 'jadi', 'bijak', 'masyarakat', 'mbg', 'indonesia', 'dpr', 'program', 'prabowo', 'perintah']
- Topic 4: ['kasus', 'jaksa', 'prabowo', 'serah', 'tambang', 'belitung', 'pt', 'bangka', 'smelter', 'timah']
- Topic 6: ['pusat', 'jadi', 'teguh', 'revitalisasi', 'bangun', 'proyek', 'wali', 'gibran', 'kota', 'solo']
- Topic 7: ['prasetyo', 'kota', 'presiden', 'ibu', 'gibran', 'papua', 'wapres', 'kantor', 'bangun', 'ikn']



Figure 5. Top words between pre and post inauguration periods

A comparison of topic similarity was then carried out by comparing the cosine similarity of the top 5 topics, with the highest score being 0.61 between topic pre_9 and topic post_0. It is useful to compare topic closeness between documents based on term of their subject [27]. The topic comparisons were done between two periods of time which are pre-inauguration and post-inauguration.

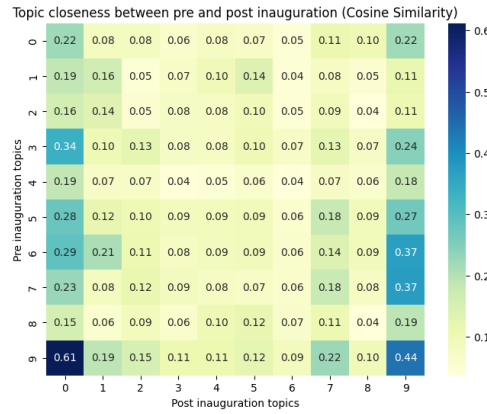


Figure 6. Cosine similarity between pre and post periods

To observe topic trends during both periods, a search was first conducted for dominant topics and compared across the two periods. From the trend analysis results, Gibran showed a strong trend in topic 4 which is related to a city of Solo, while Prabowo Subianto had the strongest trend in topic 0 after the inauguration (pasca_pelantikan). Meanwhile, the three strongest topic trends before the inauguration (pra_pelantikan) were topic 2, 5, and 0 whereas four topic trends after inauguration are topic 0, 7, 8, 9.

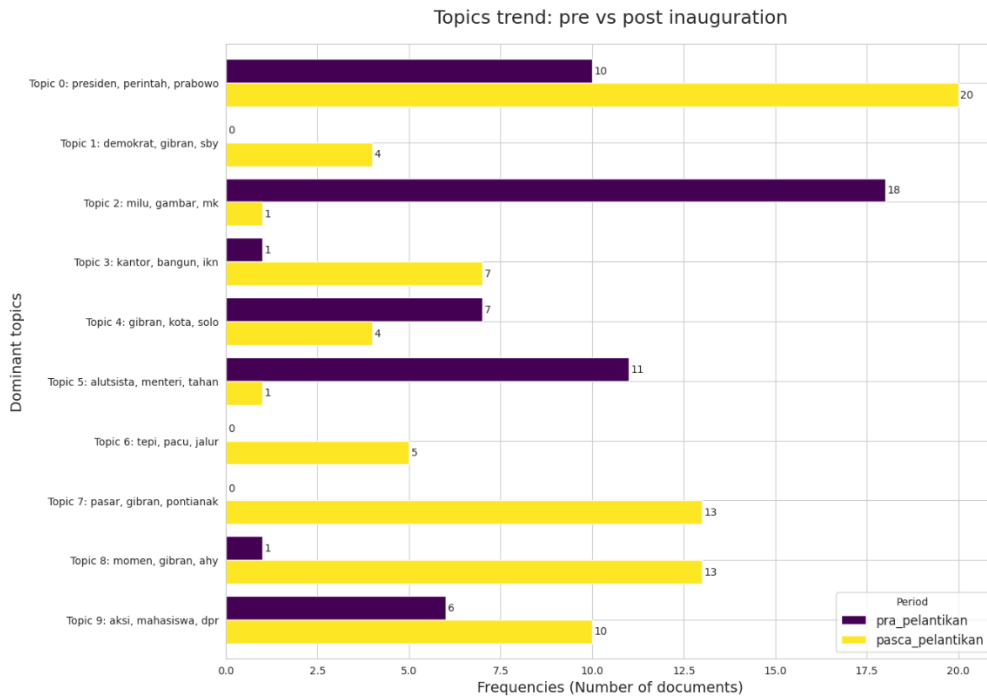


Figure 7. Topic trends between pre and post periods

4. CONCLUSION

The utilization of big data in topic models has been widely applied across various public sectors such as rhetorical analysis, political science, disaster management, including public policy using LDA. This due to the fact that the increasing number of studies in the fields of disaster management and public policy before and after the COVID-19 pandemic. Our study has succeeded implementing LDA model and compare it with NMF to extract hidden topics within documents and successfully organizing topic modelling for news topic extraction. Evaluation results implies that both models have reliable performance and possibility to be improved through an iterative refinement process. The coherence score for LDA reached 0.3709 and NMF

0.68 indicating that NMF outperformed than LDA which is producing similar result according to prior study when evaluating such algorithms like BERTopic, LDA, NMF and Top2Vec.

In the context of public policy-related, we can retrieve some keywords narrated in national media about Prabowo-Gibran administration before and after inauguration day and yet not coverage all news according to government whole programs in the dataset. Therefore, we need to recapitulate our methodology for this domain prudently.

For next roadmap, in order to refine our future data and method, we will classify online news article with machine learning model particularly Indonesian news classification event though LDA provides a good way to analyze large amount of uncategorized data. Meanwhile, the data in this study collected by pre-organizing data with keywords search like “Prabowo”, “Gibran”, and “Policy” which was not cover all news discussing about policy-related. In term of topic modelling algorithm, we will consider another unsupervised machine learning such as “top2vec” model for the future study because it is free of human interpretation and no need preprocessing step such as removing stop-words, stemming, and lemmatization of text like we did in LDA model however it is need to be develop further for Indonesia-language text. Even a recent research need to be considered by turning to topic modelling incorporated Artificial Intelligence (AI), notably Large Language Model (LLM) for designing and evaluating LDA model for specific studies objective.

REFERENCES

- [1] A. Dharma and N. Hendri, “Urgensi Penggunaan Big Data Analytics dalam Audit Sektor Publik,” vol. 18, pp. 107–120, 2022.
- [2] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, “Text mining in big data analytics,” *Big Data Cogn. Comput.*, vol. 4, no. 1, pp. 1–34, 2020.
- [3] S. Bagaskoro, I. Amrozi, and A. R. Ramadhan, “Political Sentiment Analysis,” *J. PolGov*, vol. 4, no. 2, pp. 189–230, 2022.
- [4] S. Daud, M. Ullah, A. Rehman, T. Saba, R. Damaševičius, and A. Sattar, “Topic Classification of Online News Articles Using Optimized Machine Learning Models,” *Computers*, vol. 12, no. 1, 2023.
- [5] N. Newman, A. R. Arguedas, C. T. Robertson, R. K. Nielsen, and R. Fletcher, “Reuters Institute Digital News Report 2025.” 2025.
- [6] A. F. Hidayatullah, M. R. Ma’arif, M. Habibie, and S. Khomsah, “Indonesia Infrastructure Development Topic Discovery on Online News with Latent Dirichlet Allocation,” in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1077, no. 1, p. 12012.
- [7] D. Antons, E. Grünwald, P. Cichy, and T. O. Salge, “The application of text mining methods in innovation research: current state, evolution patterns, and development priorities,” *R&D Manag.*, pp. 329–351, 2020.
- [8] P. Ghasiya and K. Okamura, “Investigating COVID-19 News across Four Nations: A Topic Modeling and Sentiment Analysis Approach,” *IEEE Access*, vol. 9, pp. 36645–36656, 2021.
- [9] E. M. S. Saleh, N. A. Rahman, and D. H. Santoso, *How Trends Shape the Media Landscape*. Universiti Pendidikan Sultan Idris.
- [10] M. A. Ramdhani, M. A. Ramdhani, D. S. A. Maylawati, and T. Mantoro, “Indonesian news classification using convolutional neural network,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 2, pp. 1000–1009, 2020.
- [11] H. S. Maghdid, “Web News Mining Using New Features: A Comparative Study,” *IEEE Access*, vol. 7, pp. 5626–5641, 2019.
- [12] L. L. Wang and K. Lo, “Text mining approaches for dealing with the rapidly expanding literature on COVID-19,” *Brief. Bioinform.*, 2021.
- [13] R. Churchill and L. Singh, “The Evolution of Topic Modeling,” *ACM Comput. Surv.*, vol. 54, no. 10, 2022.
- [14] W. H. Rochmawati and others, “Analisis Persepsi Masyarakat Terhadap Komunikasi Kebijakan Menggunakan Topic Modelling,” 2022.
- [15] R. Raditya, “Perumusan Isu Strategis Kebijakan Investasi pada Sektor Pariwisata di Indonesia Berbasis Berita Media Online,” *Multiverse Open Multidiscip. J.*, vol. 2, no. 3, pp. 346–362, 2023.
- [16] S. I. Ishak, O. Arnilia, T. Widodo, and I. G. N. A. B. Tatwa, “Analisis Sentimen terhadap Pemerintahan Prabowo–Gibran menggunakan IndoBERT dan LDA,” *Jambura J. Informatics*, vol. 1, no. 2, pp. 72–82, 2025.
- [17] W. Wahyuni, T. P. Lestari, M. Apriliana, and R. Gumelta, “Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments,” 2025.
- [18] C. Suhaeni, L. Nissa, A. Mualifah, and H. Wijayanto, “LDA Topic Modeling Analysis of Public Discourse on Indonesia’s Free Nutritious Meals Program (MBG),” *Int. J. Informatics Dev.*, vol. 14, no. 1, pp. 587–600, 2025.
- [19] D. T. Attaulah and D. Soyusiawaty, “Analisis Sentimen Program Makan Siang Gratis di Twitter/X menggunakan Metode Bi-LSTM,” *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 294–303, 2025.
- [20] D. Ayu, P. Wulandari, G. W. Wardhana, I. D. G. Puja, and W. Sindu, “Sentiment Analysis of the Free Lunch Program by Prabowo-Gibran Using the Naive Bayes Classifier,” *Sist. Kendali & Jar.*, vol. 4, 2025.
- [21] A. Farkhod, A. Abdusalomov, F. Makhmudov, and Y. I. Cho, “LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model,” *Appl. Sci.*, vol. 11, no. 23, 2021.
- [22] R. Egger and J. Yu, “A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts,” *Front. Sociol.*, vol. 7, 2022.
- [23] C. Meaney and others, “Non-negative matrix factorization temporal topic models and clinical text data identify COVID-19 pandemic effects,” *J. Biomed. Inform.*, vol. 128, 2022.
- [24] R. Debnath and R. Bardhan, “India nudges to contain COVID-19 pandemic: A reactive public policy analysis using machine-learning based topic modelling,” *PLoS One*, vol. 15, no. 9, 2020.
- [25] E. Mayor and A. Miani, “A topic models analysis of the news coverage of the Omicron variant in the United Kingdom press,” *BMC Public Health*, vol. 23, no. 1, 2023.
- [26] G. Wiedemann, *The World of Topic Modeling in R*. Nomos, 2022.
- [27] D. Gunawan, C. A. Sembiring, and M. A. Budiman, “The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents,” in *Journal of Physics: Conference Series*, 2018, vol. 978, no. 1, p. 12120.