COMPILER
Vol. 14, No. 2, November 2025, pp. 119-130
ISSN: 2252-2403 (Print), ISSN: 2549-2403 (Online)
Accredited Fourth Grade by Kemenristekdikti, Decree No:10/C/C3/DT.05.00/2025
DOI: 10.28989/compiler.v14i2.3499

Topic Analysis of Indonesian Online News on the Free Nutritious Meal Program Using Non-Negative Matrix Factorization

Irmma Dwijayanti[1,*], Alfirna Rizqi Lahitani[2], Muhammad Habibi[3]
[1]Information System, Universitas Jenderal Achmad Yani Yogyakarta
[2]Information Technology, Universitas Jenderal Achmad Yani Yogyakarta
[3]Informatics, Universitas Jenderal Achmad Yani Yogyakarta

Article Info

Article history:

Received October 3, 2025
Accepted December 23, 2025
Published December 24, 2025

Keywords:

Free Nutritious Meal Program
NMF
MBG
Online News
Topic Modeling

ABSTRACT

Information about the MBG program is widely disseminated through online news media. However, the long-text nature of news articles presents challenges such as textual complexity, mixed topics, diverse writing styles, and the need for well-formulated feature representation. Moreover, long-text corpora are still rarely utilized in text analysis, especially in the context of policy studies in Indonesia. This study applies Non-Negative Matrix Factorization (NMF) for topic modeling on a long-text corpus of 5,390 digital news articles collected from seven national portals, with the aim of mapping public discourse on MBG. The optimal number of topics was determined using the coherence score, yielding nine distinct themes. Findings indicate that media coverage primarily revolves around program distribution in schools, the role of Micro, Small, and Medium Enterprises (MSMEs) and the food sector, budget allocation, political dynamics of national figures, and health-related concerns such as student poisoning cases. The results suggest that MBG is widely perceived as a strategic policy with broad implications for public policy, economic development, political debate, and social welfare. Methodologically, this research highlights the strength of NMF in processing long-text corpora, as its non-negative matrix decomposition can clearly reveal word association patterns and easily interpreted even with a limited vocabulary, making it a reliable approach for media analysis and public opinion studies based on online news texts.

Corresponding Author:

Irmma Dwijayanti,
Information System,
Universitas Jenderal Achmad Yani Yogyakarta,
Jl. Siliwangi Jl. Ringroad Barat, Area Sawah, Banyuraden, Kec. Gamping, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55293
Email: *irmmadwijayanti@gmail.com

1. INTRODUCTION

The Free Nutritious Meal Program, or Program Makan Bergizi Gratis (MBG) in Indonesian, was recently launched by the government as an ambitious initiative to tackle malnutrition and stunting. The program is designed to provide adequate nutritional intake across various regions as part of a national strategy to enhance human resource quality, particularly for school-aged children [1]. In media and public reports, MBG has become a widely debated issue, especially regarding its implementation, sustainability, food safety, and long-term effectiveness.

In today's digital era, information about MBG is widely disseminated through online news platforms. Digital news is often considered more credible compared to informal sources on social media, as it is generally

managed by media institutions with reputations and editorial mechanisms [2]. Thus, online news corpora serve as a strategic source for accessing public debates and opinions regarding MBG.

The characteristics of online news corpora consist of long-text articles with relatively complete narrative structures, sequential paragraphs, contextual backgrounds, quotations, statistical data, and opinions. Long texts have advantages such as providing deeper context and connecting dispersed ideas across documents compared to short texts like tweets or comments [3]. However, understanding public perceptions from long texts presents challenges, as lengthy documents often involve textual complexity [4], mixed topics, greater stylistic variations [5], and the need for well-formulated feature representation. Extracting unique and coherent topics from long texts is also difficult since they usually contain multiple themes, requiring modeling techniques capable of disentangling latent topics at scale while filtering relevant information.

One of the most widely used approaches for uncovering thematic structures in text collections is topic modeling [6]. Over the past five years, numerous studies have tested and compared various topic modeling techniques on text data—including news, social media, and online forums—to reveal latent themes related to public opinion, policy issues, or media discourse. For example, [7] compared four techniques—Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), Top2Vec, and BERTopic—on Twitter data and found that NMF and BERTopic offered superior interpretability in social contexts compared to LDA and Top2Vec. Topic modeling with LDA combined with Cosine Similarity has also been applied to classify tweets for detecting mental health issues in Indonesia based on the discussed symptoms [8]. In the same year, LDA was employed to analyze topic trends in negative reviews of mobile banking applications to identify user complaints [9]. Another study combined Contextualized Topic Models (CTM) with MPNet to extract topics from user feedback with greater accuracy and contextual relevance [10].

In research [11] evaluated the performance of NMF and LDA using metrics such as coherence and cluster analysis (silhouette) to assess topic quality on short-text data, showing that NMF could produce more distinct topics in some cases, although LDA remained superior in classical probabilistic metrics. Another study compared the performance of LDA, NMF, and BERTopic on a corpus of news headlines, revealing that BERTopic slightly outperformed in coherence according to human and large language model evaluations, while NMF remained competitive and simpler to interpret [12]. Similarly, [13] employed a combination of LDA, Latent Semantic Analysis (LSA), and NMF in analyzing psychosocial survey data and concluded that each method has different strengths in capturing latent themes, with NMF tending to generate “sharper” topics when documents are relatively long and structured.

Most of the literature on topic modeling has focused on relatively short texts such as news headlines, tweets, comments, or brief media excerpts. Several studies have examined NMF on news or article corpora and compared it with other methods in terms of coherence and interpretability. The coherence score values of NMF in news text analysis and long-document corpora generally fall within the range of **0.45–0.65**, depending on corpus size, feature representation, and the number of topics evaluated; this range can therefore be used as a benchmark for assessing the quality of the topics produced in this study [12], [14]. However, specific applications of NMF-based topic modeling on long news corpora focusing on national policy programs like MBG remain limited, particularly within the Indonesian context. This study therefore, aims to apply NMF for topic modeling on a long-text corpus of Indonesian digital news articles to identify the main themes of public opinion surrounding MBG. The analysis of long-text news articles is chosen because mass media plays an important role in shaping public perceptions, providing legitimacy, and generating criticism toward government policies. By mapping the key issues raised in media discourse, this research is expected not only to contribute to the application of long-text analysis using NMF but also to provide practical benefits for policymakers in understanding public perceptions, anticipating implementation challenges, and improving the effectiveness of the MBG program in Indonesia.

2. RESEARCH METHOD

This study consists of three main stages: data collection, preprocessing, and NMF-based topic modeling. The research stages are illustrated in Figure 1.

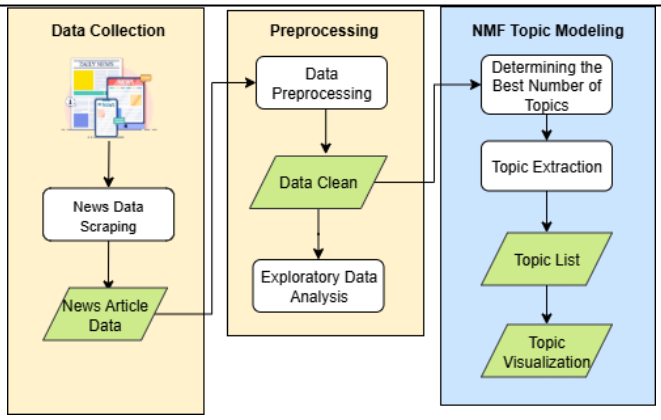

Figure 1. Research Stage

2.1. Data Collection

The data in this study were obtained through a web scraping process from Indonesian online news portals, including ANTARA News, CNBC Indonesia, detik.com, Kompas.com, MetroTV News, Okezone, and Tribun News. Web scraping is the process of extracting data from specific web pages [15]. The keyword used during data collection was *"makan bergizi gratis"*, ensuring that the retrieved articles were relevant to the research topic. A total of 5,390 news articles were collected and subsequently used as the research corpus. The distribution of articles across each news portal is presented in Figure 2.

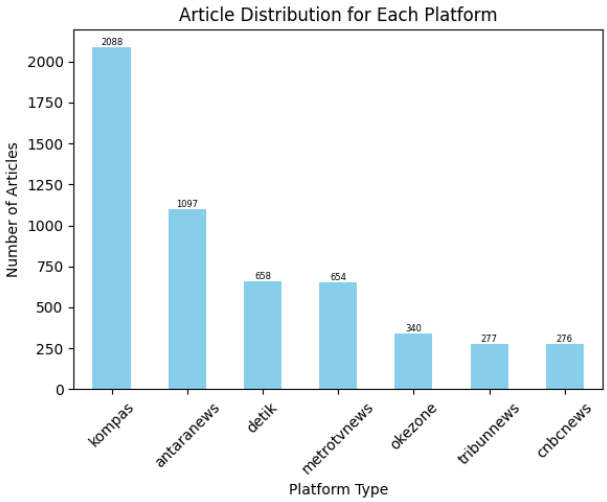

Figure 2. Distribution of the Number of Articles per Platform

2.2. Data Preprocessing

The preprocessing stage was carried out to clean the text from irrelevant elements, thereby producing data that are ready for further analysis [16]. The preprocessing steps included: *casefolding*, which converts all characters in the text into lowercase to standardize word forms; *punctuation removal*, which eliminates punctuation marks, numbers, and non-alphabetic characters; and *stopword removal*, which removes common words that frequently appear but carry little semantic meaning in the context of topics, such as "yang", "dan", and "di". After preprocessing, the number of articles available for analysis was reduced to 5,387 articles. This preprocessing procedure represents a standard practice in text mining to improve data quality prior to conducting further analysis.

2.3. Exploration Data Analysis (EDA)

The EDA stage was conducted to understand the initial characteristics of the text data before proceeding to topic modeling [17]. Based on the results of data exploration, the average number of words per article was 115, indicating that most news articles fall into the medium-length category, with a few outliers consisting of very long articles. Figure 3 presents the word count distribution per article, showing that the majority of articles range between 80 and 150 words, with the distribution peak occurring around 100–120 words. This reflects the tendency of Indonesian online media to provide short to medium-length narratives for quick reader consumption. Articles exceeding 400 words are relatively rare and may be categorized as in-depth or feature news.

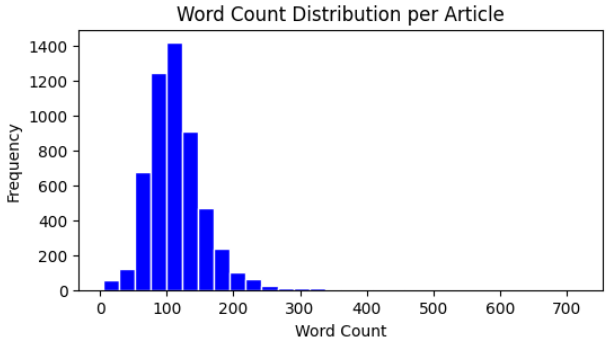

Figure 3. Word Count Distribution per Article

Figure 4 displays the most frequent words appearing in the articles. The word frequency analysis highlights the top 20 most frequent words in the corpus after preprocessing. The most frequent word was *"mbg"*, followed by *"makan"*, *"bergizi"*, and *"gizi*. Additionally, terms related to policy and program implementation also appeared prominently, such as *"pemerintah"*, *"anggaran"*, *"penerima"*, *"pelaksanaan"*, as well as political actors like *"prabowo"*. This distribution indicates that the corpus consistently discusses MBG within the context of government policy, budgeting, target beneficiaries, and political dynamics.

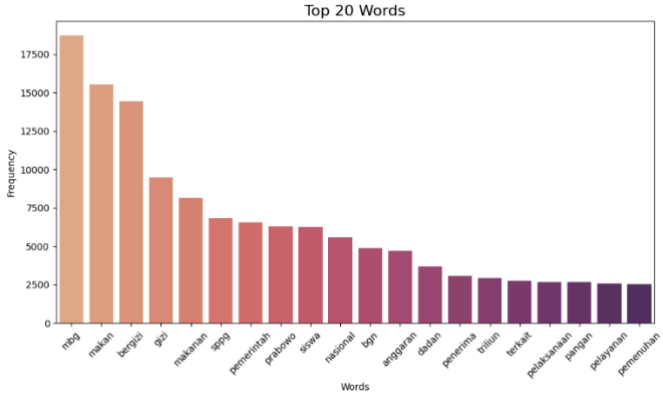

Figure 4. Top Words

Overall, the EDA results demonstrate that the corpus is primarily dominated by themes relevant to the research focus, namely, the MBG program, while also exhibiting reasonable variation in article length for topic modeling purposes. The analysis further confirms that preprocessing successfully removed non-meaningful words, allowing the dominant terms to be more representative of the study's context.

2.4. NMF Topic Modeling

Topic modeling in this study was conducted using the NMF method. NMF works by decomposing the document-term matrix (constructed from text representation using Term Frequency–Inverse Document Frequency, TF-IDF) into two non-negative matrices, namely the document–topic matrix (W) and the topic–term matrix (H) [18]. Equation (1) illustrates how the document–term matrix V is factorized into the two non-negative matrices W and H [19].

$$V \approx W \, x \, H \tag{1}$$

Because all weights are non-negative, the results of the decomposition tend to be more interpretable: each topic emerges as an additive combination of dominant words, without the "negative" values that often complicate interpretation. Compared to probabilistic methods such as Latent Dirichlet Allocation (LDA), NMF offers advantages when dealing with high-dimensional data or long-text corpora such as news articles, as it can generate sharper and more interpretable topics through linear algebraic patterns [7]. The steps in NMF-based topic modeling include: (a) determining the optimal number of topics (k) by comparing coherence scores across different topic numbers; (b) extracting topics by selecting the highest-weighted terms within each topic; and (c) generating a list of topics that includes the dominant words for each. NMF was chosen because it can produce more interpretable topic representations compared to other methods, such as LDA, particularly when applied to high-dimensional text data.

3. RESULTS AND ANALYSIS

3.1. NMF Topic Modeling Across All News Portals

The topic modeling stage was carried out using the NMF method with feature representation based on TF-IDF. The optimal number of topics was determined by evaluating the coherence score across several variations in the number of topics. The coherence score is used to measure the level of semantic relatedness among dominant words within a topic based on their co-occurrence in the corpus, such that higher values indicate topics that are more meaningful and easier to interpret. This metric has been widely used as a primary indicator of topic quality in NMF, LDA, LSA, and other topic modeling methods, as it has been shown to correlate well with human judgments of topic quality [20].

The range of topic numbers was tested from k = 2 to k = 12 with a one-topic interval (step = 1). The lower bound of k = 2 was chosen to avoid overly simplistic models that fail to capture the thematic diversity of the corpus, while the upper bound of k = 12 was set to prevent excessive topic fragmentation and to maintain result interpretability, considering the size and complexity of the analyzed data. Testing with a one-topic interval was conducted to obtain a more granular evaluation of changes in coherence values. The results show that the highest coherence score was achieved at k = 9 with a score of 0.6437, leading to the selection of a nine-topic model as the optimal solution, as illustrated in Figure 5.

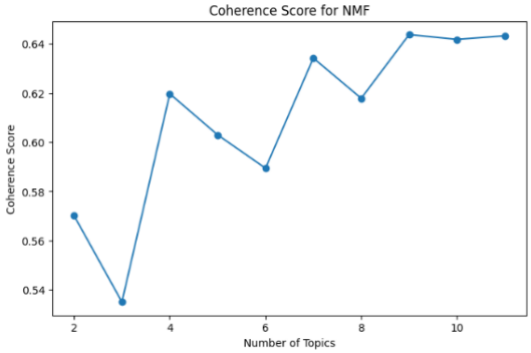

Figure 5. NMF Coherence Score Overall Portal

The list of topics generated, along with the top 20 contributing terms for each topic, is presented in Table 1. Each topic was interpreted based on the cluster of dominant terms. This table provides an overview of the key issues emerging in MBG-related news coverage, ranging from aspects of program implementation and budgeting to political dynamics and health issues. Through this interpretation, the main foci of media framing on MBG, as well as thematic variations across topic groups, can be identified.

Table 1. List of Topic

Topic No.	Topic-Forming Keyword	Topic Interpretation
1	'sppg', 'gizi', 'mbg', 'penerima', 'pelayanan', 'bgn', 'pemenuhan', 'satuan', 'polri', 'beroperasi', 'pembangunan', 'melayani', 'target', 'polda', 'kecamatan', 'hasan', 'sigit', 'nasional', 'menyusui', 'kapolri',	The Role of Public Officials and Public Services
2	'triliun', 'anggaran', 'penerima', 'apbn', 'mulyani', 'tambahan', 'sri', 'mencapai', 'keuangan', 'mbg', 'pemerintah', 'alokasi', 'efisiensi', 'pendidikan', 'zulhas', 'kementerian', 'pendapatan', 'miliar', 'rapbn', 'target'	Financing and Fiscal Policy of the MBG Program
3	'dadan', 'bgn', 'hindayana', 'nasional', 'penerima', 'mbg', 'gizi', ['gizi', 'makan', 'bergizi', 'makanan', 'pelayanan', 'satuan', 'melayani', 'rapat', 'telur', 'terkait', 'target', 'istana', 'dibawa'	Coordination of the Central Government
4	'keracunan', 'siswa', 'mengalami', 'cianjur', 'kejadian', 'makanan', 'mbg', 'gejala', 'kesehatan', 'diduga', 'sampel', 'menyantap', 'mual', 'penyebab', 'sukoharjo', 'bpom', 'dirawat', 'laboratorium', 'dinkes', 'mengonsumsi'	Poisoning Cases and Health Issues
5	'siswa', 'makanan', 'makan', 'bergizi', 'mbg', 'sdn', 'gizi', 'sayur', 'gibran', 'pendidikan', 'disajikan', 'pelaksanaan', 'kecamatan', 'dibagikan', 'tk', 'pelajar', 'sehat', 'menerima', 'smpn', 'mengaku'	Distribution of the MBG Program Directly in Schools
6	'umkm', 'pangan', 'mbg', 'pemerintah', 'koperasi', 'gizi', 'mendukung', 'kerja', 'nasional', 'pembangunan', 'meningkatkan', 'pertumbuhan', 'sektor', 'sapi', 'kesehatan', 'pelaku', 'kementerian', 'generasi', 'mendorong'	The Role of MSMEs, Cooperatives, and the Food Sector in Supporting the Program
7	'yayasan', 'ira', 'kalibata', 'mbn', 'metro', 'danna', 'harly', 'penggelapan', 'mbg', 'mesra', 'pembayaran', 'dugaan', 'laporan', 'bgn', 'timoty', 'nusantara', 'pancoran', 'kliennya', 'miliar', 'melaporkan'	Legal Issues and Alleged Irregularities
8	'prabowo', 'makan', 'gates', 'subianto', 'bergizi', 'bill', 'hasan', 'prabowo', 'sidang', 'istana', 'kabinet', 'gibran', 'jati', 'mbg', 'meninjau', 'rights', 'reserved', 'pemerintahan', 'kepresidenan', 'pelaksanaan', 'brasil'	Political Narratives, National Figures, and Leadership Dynamics Related to MBG

9	'zakat', 'usulan', 'makan', 'dpd', 'pendanaan', 'bergizi', 'membiayai', 'pemerintah', 'ketua', 'fakir', 'infak', 'miskin', 'mengusulkan', 'anggaran', 'golongan', 'menstimulus', 'najamudin', 'dna', 'dipakai', 'berpikir'	Alternative Financing for MBG through Zakat, Infaq, or Other Social Funds

Figure 6 illustrates the distribution of articles across each topic resulting from the NMF model. It can be observed that Topic 5 (Distribution of MBG in schools) and Topic 6 (Role of MSMEs & local food sector) dominate with the largest number of articles. This indicates that media coverage tends to emphasize the program's implementation in schools alongside the involvement of MSMEs and the food sector as key elements of program sustainability. Other topics with relatively high article counts include Topic 8 (Political narratives & national figures), Topic 2 (Financing and fiscal policy), and Topic 1 (Public services & officials). These findings highlight that issues of budgeting, government policy, and political actors are central to media discourse on MBG.

In contrast, the topics with the fewest articles were Topic 7 (Legal Issues and Alleged Irregularities) and Topic 9 (Alternative financing for MBG). The relatively small volume of coverage suggests that legal matters and alternative funding schemes are not yet mainstream in media narratives. Overall, the distribution confirms that the media tends to emphasize technical implementation, distribution, MSME involvement, and national politics and policy, rather than peripheral issues such as alternative financing or allegations of irregularities.

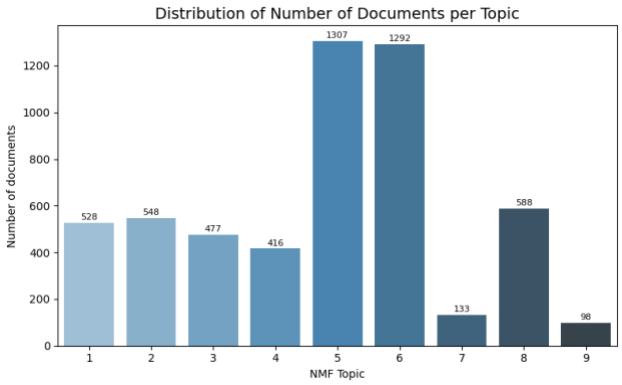

Figure 6. Distribution of Number of Documents per Topic

3.2. NMF Topic Modeling by Individiual News Portal

Topic modeling for each news portal also employed the coherence score to determine the most appropriate number of topics. The coherence score was calculated by testing topic numbers ranging from 2 to 10. Figure 7 presents the coherence score results for each portal, showing that every outlet produced a different optimal number of topics. These variations are influenced not only by the diversity of issues within the texts but also by the size of the sample-feature space (i.e., the combination of documents and vocabulary formed after preprocessing for each portal) [21].

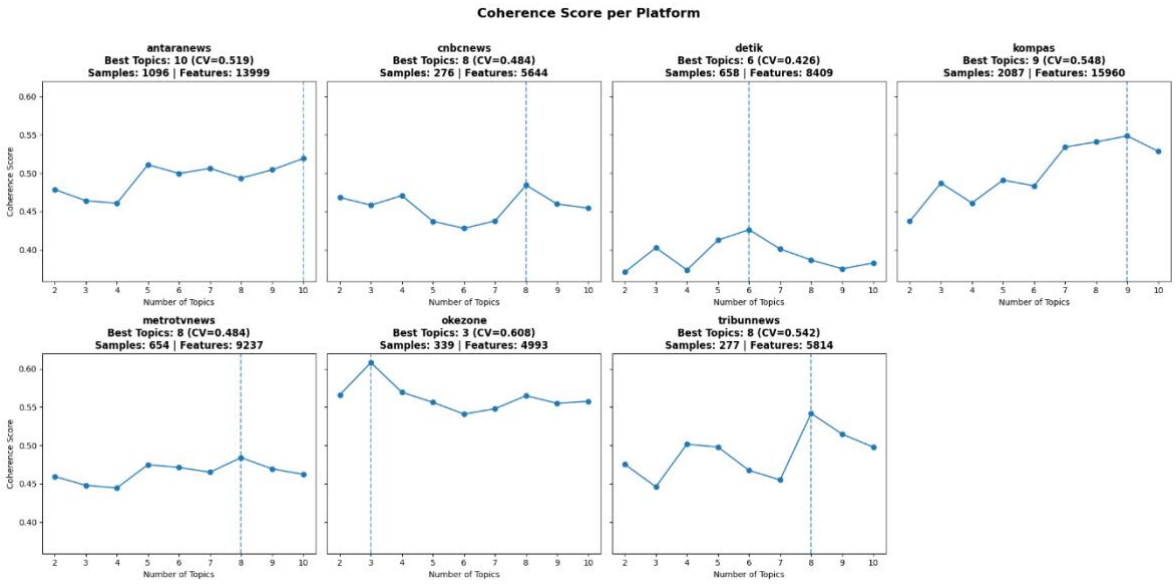

Figure 7. Coherence Score per Portal

In general, a larger sample-feature size allows for richer vocabulary variation that can be grouped into more topics. For example, Kompas, which had the largest sample-feature set, produced 9 topics, while ANTARA News, with 1,096 sample-features, generated 10 topics, reflecting broad issue coverage. Interestingly, smaller sample-feature sizes do not always correspond to fewer topics. For instance, Tribunnews had only 277 sample-features yet produced 8 topics, indicating that the NMF algorithm can still cluster diverse issues even with limited vocabulary. This suggests that NMF does not rely solely on vocabulary size but also on the distribution and strength of word associations within the corpus. This finding is consistent with [22], which highlights that NMF topic coherence is shaped more by latent word association structures than by feature counts alone, and with studies on topic modeling stability through matrix decomposition [23]. Thus, even with limited features, NMF can still generate rich and meaningful topics when strong contextual variation is present.

After determining the optimal number of topics, NMF modeling was applied to each portal, and the results were visualized using wordclouds. Figure 8 shows the wordcloud displaying the distribution of dominant words in the ANTARA News portal. ANTARA News presents wide issue coverage with ten topics, including combinations of words such as *anggaran, Pembangunan, sppg, pelayanan, keracunan, polri, pangan, umkm.* This portal tends to emphasize official narratives and national policies in line with its role as a government news agency, with reporting focusing on MBG implementation in the context of national development, state institutional support (e.g., police, ministries), and macro issues such as food security and public services. Health issues are also highlighted, such as student poisoning cases, but remain linked to regulations and government interventions. Thus, ANTARA's framing reflects MBG as part of the national development agenda integrated with other sectors.

Figure 8. Antaranews Portal Wordcloud

Figure 9 shows the wordcloud of dominant words in the CNBC Indonesia portal. CNBC Indonesia frames MBG in terms of economics, investment, and state finance, consistent with its media orientation. The wordcloud highlights dominant words such as *triliun, APBN, mulyani, sektor, agro, tbk, investasi, surplus,* and *burden sharing*. This indicates a focus on fiscal and macroeconomic implications, including MBG financing sustainability in the RAPBN, the role of the food industry (e.g., Indofood, Nusantara Agro), and national political economy narratives. CNBC also discusses the involvement of corporations, MSMEs, and cooperatives in supporting the program, positioning MBG not only as a social policy but also as a strategic economic instrument related to growth, investment, and food markets. These findings are consistent with [7], which compared LDA and NMF in news text analysis, showing that NMF is better at highlighting coherent themes on policy and economic issues in news corpora.

Figure 9. CNBC Indonesia Portak Wordcloud

Figure 10 shows the wordcloud of dominant words in the detikcom portal. Detikcom generated six topics that are relatively focused on field implementation and the practical impacts of MBG. Dominant words such as *pelaksanaan, mbg, sppg, siswa, makanan, keracunan,* and *distribusi* indicate attention to issues close to the public. Additionally, topics such as *anggaran, APBN,* and *penerima* reinforce the link between policy and field realities. Detik also highlights health issues, particularly student poisoning cases, reflecting a focus on popular public issues. The correlation between Detik.com's coverage of poisoning and [24] highlights that food poisoning incidents in MBG reflect gaps in oversight and the state's technical capacity to ensure food safety. Thus, Detikcom's framing tends to be populist and oriented toward everyday issues directly experienced by society, with touches of policy and political dynamics.

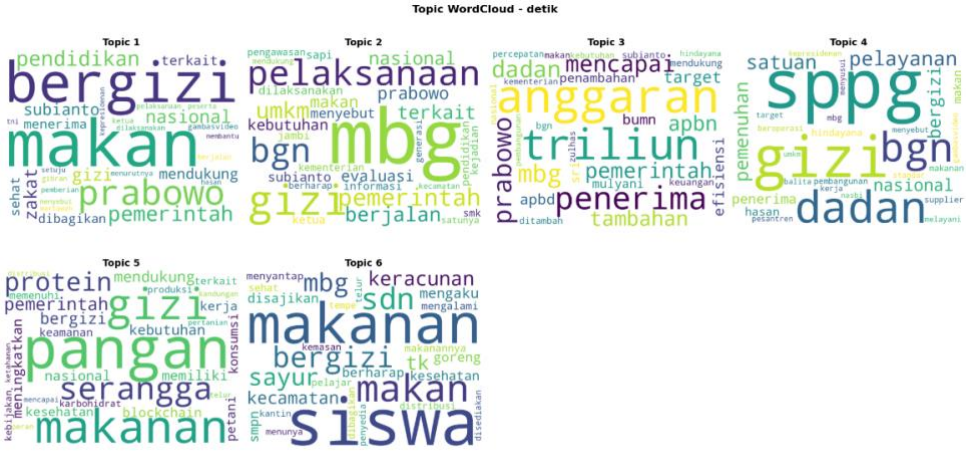

Figure 10. Detikcom Portal Wordcloud

Figure 11 shows the wordcloud of dominant words in the Kompas portal. Kompas presents nine topics with a balance between political, policy, and social issues. The wordcloud shows keywords such as *anggaran, efisiensi, prabowo, pemerintah, bgn, keracunan, yayasan,* and *pelaksanaan*. These topics indicate that Kompas not only covers the technical implementation of MBG in schools but also highlights political controversies (national figures), health issues (student poisoning), and institutional governance (allegations of foundation misuse). Kompas' reporting tends to be comprehensive and investigative, framing MBG as a multidimensional policy involving politics, economy, society, and potential governance challenges at the grassroots level.

Figure 11. Kompas Portal Wordcloud

Figure 12 shows the wordcloud of dominant words in the MetroTV News portal. MetroTV News generated eight topics with an emphasis on government policy narratives and national political dynamics. The wordcloud highlights words such as *pemerintah, pelaksanaan, prabowo, pendidikan, keracunan, koperasi,* and *pangan*. This confirms that MetroTV frames MBG as part of government agendas and development strategies. Technical issues such as student poisoning are also covered, though more as illustrations of policy impact. Political topics (e.g., Prabowo, cabinet, presidency) further show that MetroTV's framing is inseparable from the dimensions of politics and government communication, positioning MBG as a policy to be understood within the broader framework of national stability. This framing is consistent with government policy framing studies such as [25], which demonstrate that media often associate social policy issues with institutional legitimacy and government narratives to shape public perceptions of program effectiveness.

Figure 12. MetroTV News Portal Wordcloud

Figure 13 shows the wordcloud of dominant words in the Okezone portal. Okezone generated only three topics, indicating relatively narrow and simple coverage compared to other portals. Dominant keywords such as *makan, bergizi, prabowo, mbg, sppg,* and *yayasan* suggest that Okezone focuses more on technical implementation, political figures, and local institutional issues. This framing reflects a popular news approach with limited scope, leaning toward issues that quickly capture public attention, such as school meal distribution,

involvement of national figures, and the role of foundations in program implementation. Thus, Okezone presents MBG as a practical issue, linked to public figures and local narratives.

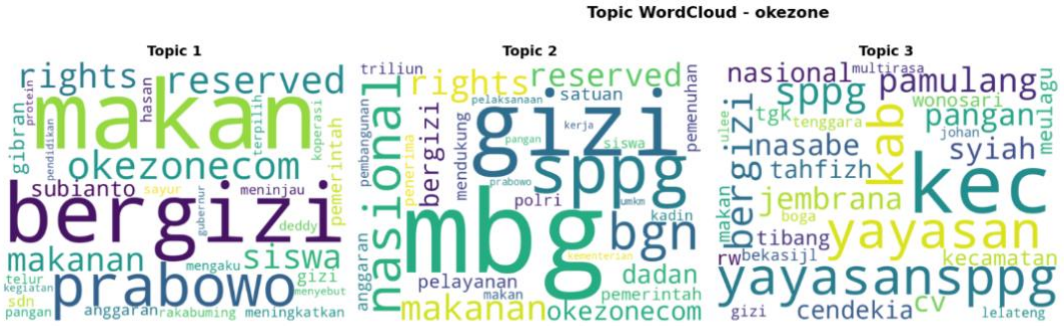

Figure 13. Okezone Portal Wordcloud

Figure 14 shows the wordcloud of dominant words in the Tribunnews portal. Tribunnews generated eight topics with broad coverage, combining political, social, and controversial issues. The wordcloud highlights words such as *makan, bergizi, gibran, prabowo, anggaran, keracunan, yayasan, pesantren*. This indicates that Tribunnews frames MBG through popular narratives, including political figures (Prabowo, Gibran), state budget issues, health cases (student poisoning), and socio-religious issues (pesantren, NU, santri). With its characteristics as a mass-audience news portal, Tribunnews adopts a more populist framing, prioritizing issues that are easily viral and diverse to capture reader interest across various segments. This is consistent with research on online media strategies that select sensational or viral issues as part of "viral journalism" to expand reach and audience engagement [26]. In addition, studies covering Tribunnews show that the portal often places national figures at the center of reporting within critical or contestative frames, reflecting a strategy to attract a broad audience segment through the reinforcement of popular narratives [27].

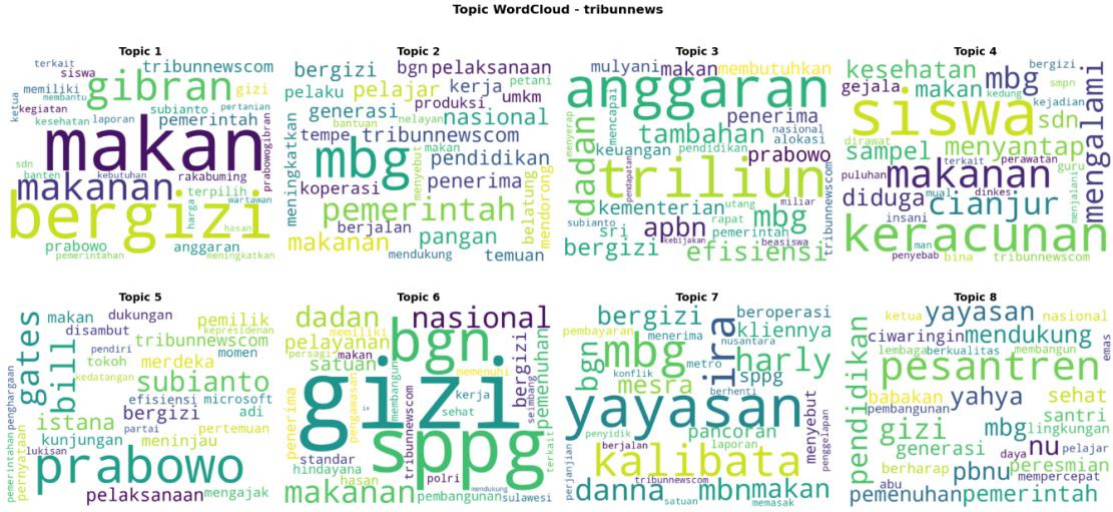

Figure 14. Tribunnews Portal Wordcloud

3.3. Discussion

Overall, the results of topic modeling and wordcloud visualizations across different news portals reveal common core focuses, namely budgeting and sustainability of MBG funding, technical implementation in schools, political narratives of national figures, and health issues such as student poisoning. Each portal, however, emphasizes different aspects in line with its orientation such as CNBC on economic-fiscal aspects, Kompas on multidimensional political-social issues, MetroTV on government policy, Okezone with a popular approach, and Tribunnews with populist framing. Nevertheless, the common thread remains that MBG is perceived as a strategic program with broad implications spanning public policy, politics, the national economy, and practical societal issues. These results are consistent with [28], which also used NMF to analyze media framing of MBG and found that news reporting tends to frame MBG through the lens of national policy, public responses, and political dynamics.

The analysis results indicate that NMF can be applied for topic modeling on news texts with varying article lengths, including relatively long articles, such as MBG coverage across multiple media portals. The

topics generated by NMF tend to be more clearly separated and semantically consistent, particularly in capturing key issues such as budget policy, program implementation, political actors, and health-related cases. These findings are consistent with previous studies reporting that NMF is capable of producing stable and easily interpretable topics on news corpora and long documents [18].

From an evaluation perspective, the use of the coherence score as the criterion for selecting the optimal topic model strengthens the validity of the NMF results, as this metric directly assesses the semantic relatedness among dominant words within each topic. In this study, the relatively high coherence score of 0.6437 indicates that the generated topics are not only mathematically well separated but also contextually relevant to media framing. This coherence-based evaluation approach is also widely applied in LDA studies on long corpora, and prior research shows that both LDA and NMF can produce meaningful topics when the number of topics is determined based on peak coherence values, thereby supporting the use of NMF as a reliable and appropriate method for cross-media topic analysis of MBG news [14].

4. CONCLUSION

This study applied the NMF method for topic modeling on a long-text corpus of 5,390 Indonesian digital news articles to identify dominant themes in public discourse regarding the MBG. The findings show that recurring issues across news portals include budget and funding sustainability, technical implementation in schools, political narratives involving national figures, and health concerns such as student poisoning. While framing variations reflect each media outlet's editorial orientation, MBG is generally perceived as a strategic program with broad implications for public policy, the national economy, political dynamics, and social issues. From a methodological perspective, NMF was able to generate coherent and interpretable topics despite variations in sample-feature sizes across portals. Its non-negative matrix decomposition effectively reveals word association patterns, making it a reliable method for analyzing long-text corpora. These results highlight the potential of NMF for media analysis and public opinion research based on digital news texts, particularly in examining complex and multidimensional policy programs such as MBG.

REFERENCES

[1] Inspektorat Jenderal Kemendikbudristek, "Program Makan Bergizi Gratis: Langkah Strategis Pemenuhan Gizi dan Penguatan Karakter Anak Bangsa," https://itjen.kemdikbud.go.id/. Accessed: Feb. 02, 2025. [Online]. Available: https://itjen.kemdikbud.go.id/web/program-makan-bergizi-gratis-langkah-strategis-pemenuhan-gizi-dan-penguatan-karakter-anak-bangsa/

[2] A. Ceron, "Internet, News, and Political Trust: The Difference between Social Media and Online Media Outlets," *J. Comput. Commun.*, vol. 20, no. 5, pp. 487–503, Sep. 2015, doi: 10.1111/JCC4.12129.

[3] P. Saha, A. Kevin, and C.-C. Chang, "Enhancing Short-Text Topic Modeling with LLM-Driven Context Expansion and Prefix-Tuned VAEs," Oct. 2024, Accessed: Oct. 02, 2025. [Online]. Available: https://arxiv.org/pdf/2410.03071

[4] E. Puspita, F. Shiddieq, and F. F. Roji, "Pemodelan Topik pada Media Berita Online Menggunakan Latent Dirichlet Allocation (Studi Kasus Merek Somethinc)," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 481–489, Feb. 2024, doi: 10.57152/MALCOM.V4I2.1204.

[5] A. Goyal and I. Kashyap, "Comprehensive Analysis of Topic Models for Short and Long Text Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 12, pp. 249–259, Dec. 2023, doi: 10.14569/IJACSA.2023.0141226.

[6] M. Khan, S. S. Khan, and Y. Alharbi, "Text Mining Challenges and Applications, A Comprehensive Review," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 20, no. 12, pp. 138–148, 2020, Accessed: Oct. 01, 2025. [Online]. Available: https://www.researchgate.net/publication/348409623_Text_Mining_Challenges_and_Applications_A_Comprehensive_Review?

[7] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Front. Sociol.*, vol. 7, p. 886498, May 2022, doi: 10.3389/FSOC.2022.886498/BIBTEX.

[8] I. Dwijayanti, M. Habibi, K. Kusumaningtyas, and S. Riyadi, "Tweets Classification of Mental Health Disorder in Indonesia Using LDA and Cosine Similarity," *Telemat. J. Inform. dan Teknol. Inf.*, vol. 21, no. 1, pp. 41–54, 2024, doi: 10.31515/telematika.v21i1.10725.

[9] K. Kusumaningtyas, I. Dwijayanti, A. R. Lahitani, and M. Habibi, "Analisis Tren Topik dalam Ulasan Negatif Aplikasi M-Banking Menggunakan Latent Dirichlet Allocation," *J. FASILKOM*, vol. 14, no. 3, pp. 549–555, Nov. 2024, doi: 10.37859/JF.V14I3.8035.

[10] M. H. Asnawi, A. A. Pravitasari, T. Herawan, and T. Hendrawati, "The Combination of Contextualized Topic Model and MPNet for User Feedback Topic Modeling," *IEEE Access*, vol. 11, pp. 130272–130286, 2023, doi: 10.1109/ACCESS.2023.3332644.

[11] R. Muthusami, N. Mani Kandan, K. Saritha, B. Narenthiran, N. Nagaprasad, and K. Ramaswamy, "Investigating topic modeling techniques through evaluation of topics discovered in short texts data across diverse domains," *Sci. Rep.*, vol. 14, no. 1, pp. 1–13, Dec. 2024, doi: 10.1038/S41598-024-61738-4;SUBJMETA.

[12] O. Babalola, B. Ojokoh, and O. Boyinbode, "Comprehensive Evaluation of LDA, NMF, and BERTopic's Performance on News Headline Topic Modeling," *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 268–289, Nov. 2024, doi: 10.62411/JCTA.11635.

[13] A. Rkia, A. Fatima-Azzahrae, A. Mehdi, and L. Lily, "NLP and Topic Modeling with LDA, LSA, and NMF for Monitoring Psychosocial Well-being in Monthly Surveys," *Procedia Comput. Sci.*, vol. 251, pp. 398–405, Jan. 2024, doi: 10.1016/J.PROCS.2024.11.126.

[14] C. B. Pavithra and J. Savitha, "Topic Modeling for Evolving Textual Data Using LDA, HDP, NMF, BERTOPIC, and DTM With a Focus on Research Papers," *J. Technol. Informatics*, vol. 5, no. 2, pp. 53–63, Apr. 2024, doi: 10.37802/JOTI.V5I2.618.

[15] I. Made, A. Santosa, L. Putu, A. Prapitasari, and I. P. Ramayasa, "Sentiment Analysis on the Centralized Isolation Policy for Covid-19 Response in Bali Province," *Compiler*, vol. 11, no. 2, pp. 45–50, Dec. 2022, doi: 10.28989/compiler.v11i2.1356.

[16] A. W. Murdiyanto and M. Habibi, "Analysis of Deep Learning Approach Based on Convolution Neural Network (CNN) for Classification of Web Page Title and Description Text," *Compiler*, vol. 11, no. 2, pp. 51–58, Dec. 2022, doi: 10.28989/COMPILER.V11I2.1327.

[17] Ismail Olaniyi Muraina, Olayemi Muyideen Adesanya, Moses Adeolu Agoi, and Solomon Onen Abam, "The Necessity of Exploratory Data Analysis: How are preprocessing activities beneficial to Data Analysts and Professional Researchers in Academia? ," | *Int. J. Sci. Res. Comput. Sci. Eng.*, vol. 11, no. 3, pp. 22–28, Apr. 2023, Accessed: Oct. 03, 2025. [Online]. Available: https://ijsrcse.isroset.org/index.php/j/article/view/528

[18] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Front. Sociol.*, vol. 7, p. 886498, May 2022, doi: 10.3389/FSOC.2022.886498.

[19] A. Alfajri, D. Richasdy, and M. A. Bijaksana, "Topic Modelling Using Non-Negative Matrix Factorization (NMF) for Telkom University Entry Selection from Instagram Comments," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 485–492, 2022, doi: 10.47065/josyc.v3i4.2212.

[20] C. Sievert and K. E. Shirley, "LDAvis: A method for visualizing and interpreting topics," Association for Computational Linguistics (ACL), Jun. 2014, pp. 63–70. doi: 10.3115/V1/W14-3110.

[21] R. P. F. Afidh and Syahrial, "Pemodelan Topik Menggunakan n-Gram dan Non-negative Matrix Factorization," *J. Inf. dan Teknol.*, pp. 265–275, Sep. 2023, doi: 10.60083/JIDT.V5I1.385.

[22] Z. A. Guven, B. Diri, and T. Cakaloglu, "Evaluation of Non-Negative Matrix Factorization and n-stage Latent Dirichlet Allocation for Emotion Analysis in Turkish Tweets," *Proc. - 2019 Innov. Intell. Syst. Appl. Conf. ASYU 2019*, Sep. 2021, doi: 10.1109/ASYU48272.2019.8946435.

[23] M. Belford, B. Mac Namee, and D. Greene, "Stability of topic modeling via matrix factorization," *Expert Syst. Appl.*, vol. 91, pp. 159–169, Jan. 2018, doi: 10.1016/J.ESWA.2017.08.047.

[24] O. Sumantri Riyanto and M. Rianita Elfrida Sinaga, "Penegakan Hak Anak Atas Makanan Aman dan Sehat: Studi Kasus Keracunan Dalam Program Makan Bergizi Gratis Ditinjau dari Tanggung Jawab Negara," *Juris Humanit. J. Ris. dan Kaji. Huk. Hak Asasi Mns.*, vol. 4, no. 1, pp. 1–11, Jun. 2025, doi: 10.37631/JRKHM.V4I1.84.

[25] A. Sofian and D. N. Lestarini, "Analisis framing pemberitaan tentang kebijakan pemerintah dalam menangani kasus Covid-19," *COMMICAST*, vol. 2, no. 1, pp. 58–70, Feb. 2021, doi: 10.12928/COMMICAST.V2I1.3150.

[26] A. Denisova, "Viral journalism. Strategy, tactics and limitations of the fast spread of content on social media: Case study of the United Kingdom quality publications," *Journalism*, vol. 24, no. 9, pp. 1919–1937, Sep. 2023, doi: 10.1177/14648849221077749.

[27] A. Al Aslah, A. Anggana, and R. R. Ramadhan, "Analisis Framing dan Sentimen Media Daring terhadap Isu Dinasti Politik Keluarga Jokowi," *J. Ilmu Polit. dan Komun.*, vol. 14, no. 2, pp. 1–16, Dec. 2024, doi: 10.34010/JIPSI.V14I2.14065.