News Topic Analysis of Nusantara Capital City using Frobenius Norm and Non-negative Matrix Factorization

Luh Gede Surya Kartika[1], Komang Rinartha[2], Anggara Putu Dharma Putra[4], I Gst Ngr Pertu Agung[5]

[1]Dept. of Informatics, Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar, Indonesia, suryakartika@uhnsugriwa.ac.id

[2] Dept. of Computer System, STIKOM Bali, Indonesia, komangrinartha@gmail.com

[3]Dept. of Hindu Communication, Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar, Indonesia, ankfdd@uhnsugriwa.ac.id

[4]Dept. of Hindu Communication, Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar, Indonesia,

ABSTRACT

This research leverages Non-negative Matrix Factorization (NMF) with the Frobenius norm to analyze news articles from Kompas about the relocation of Indonesia's capital to Nusantara. The study is significant as it provides insights into public and media perceptions documented by Kompas, identifies critical issues surrounding this transformative national project, and demonstrates the utility of NMF in analyzing Indonesian-language news texts, particularly in the context of public policy and media discourse. A dataset of news articles related to Ibu Kota Nusantara was preprocessed through cleaning, normalization, stemming/lemmatization, and tokenization to prepare it for topic modeling. Using TF-IDF for feature extraction, Non-Negative Matrix Factorization (NMF) with Frobenius norm as the loss function was applied to generate topics, which were evaluated based on coherence scores and manual analysis for relevance and interpretability. This study identified five distinct topics related to Ibu Kota Nusantara (IKN) from Kompas news articles during January-March 2024, covering community preparations, toll road developments, buffer zone status, groundbreaking events, and ASN housing. Using the NMF model and c_uci coherence scoring, the study achieved a high coherence score of 0.991, indicating semantically connected terms that facilitate topic interpretation. The alignment between Wordcloud and NMF results demonstrates both methods' focus on significant terms, with Wordcloud highlighting key words and NMF providing a deeper structural analysis of topic interrelations.

Keywords (12 Bold) – Nusantara, IKN, NMF, Kompas

INTRODUCTION

The Indonesian government has decided to relocate the nation’s capital from Jakarta to Nusantara in East Kalimantan. This strategic move aims to alleviate Jakarta's growing population, environmental, and infrastructural burdens. Jakarta faces chronic issues, including severe traffic congestion, air pollution, and flood risks due to significant land subsidence. The capital relocation is a national project with profound economic, social, and cultural implications. As a major national endeavour, it has attracted widespread attention in the media, including prominent news platforms such as Kompas. Media documentation is crucial as a public record of developments and perspectives surrounding Nusantara’s establishment, providing valuable insights into public opinion, key issues, and sentiments regarding this relocation.

The extensive rise in online news consumption has coincided with an increasing interest in employing natural language processing (NLP) and machine learning for the automated analysis of news articles. These technologies aim at several elements of the news media ecosystem, enhancing end-consumers' access to news in a more efficient and tailored manner (Li et al., 2011; Vossen et al., n.d.) To encourage news generation and distribution (Clerwall, 2014). Other scholars have connected journalism and NLP, such as in news summarization (Astuti et al., 2024), news recommender (Nurjayanto & Baizal, 2024) and sentiment analysis of news (Samuels & Mcgonical, 2019).

Previous research has explored the role of media coverage in shaping public perception of Indonesia's capital relocation project, Nusantara. For instance, (Amelia, 2023) analyzed sentiment in online news and social media posts, highlighting mixed public reactions driven by concerns over environmental impacts and economic opportunities. Similarly, (Dhamayanti et al., 2024) examined news framing related to Nusantara, finding that political and economic issues dominated the narratives, while social and environmental aspects were underrepresented. Another study by (Ambarwati Puspitasari et al., 2023)utilized linguistic corpus analysis to trace trends and patterns in the usage of keywords related to Nusantara across media platforms, revealing shared concerns about infrastructure readiness and governance. These studies emphasize the importance of media in influencing public discourse on major national projects. Building on this foundation, the current research applies topic modeling to systematically identify themes and shifts in the coverage of Nusantara in Kompas, providing a comprehensive analysis of media framing on this significant national endeavor.

Topic modeling is a text analysis technique designed to identify hidden patterns or themes within a collection of documents (Sutherland et al., 2020). This technique is highly useful for categorizing documents based on their underlying topics without requiring manual annotation (Lotto et al., 2023). Popular methods for topic modeling include Latent Dirichlet Allocation (LDA) (Mifrah, 2020a; Pan et al., 2015), Probabilistic Latent Semantic Analysis (PLSA)(Steyvers & Griffiths, 2007) , and Non-negative Matrix Factorization (NMF) (Mifrah, 2020a). Topic modeling is widely applied to extract trends from large-scale text data, such as news articles, social media, and academic papers. By employing topic modeling, it becomes more effective and efficient to analyze data on a large scale.

Non-negative Matrix Factorization (NMF) is a topic modeling technique that decomposes a document matrix into two factor matrices representing topic and word weights (Alfajri et al., 2022). NMF has an interpretability advantage due to its non-negativity constraint, which is more intuitive for representing word frequencies or topic weights. In topic modeling, NMF works by decomposing the document-word matrix into two factor matrices (Mifrah, 2020b)—one representing word-topic relations and the other representing document-topic relations, each depicting documents and words in terms of topics (Athukorala & Mohotti, 2022). In the context of news documents, NMF enables the identification of dominant news themes related to the keywords, which can be connected to specific issues such as environmental, social, or political topics. The Frobenius norm is a standard measurement method used in matrix decomposition, including in NMF. Frobenius norm calculates the difference between the original matrix and the estimated factor matrices in NMF. It works by computing the difference between the original and the predicted matrices, squaring these differences to yield an error value. This error value is then minimized to reduce discrepancies, ensuring that the topic representations derived from NMF closely

approximate the original document matrix. Using the Frobenius norm in NMF aids in improving model accuracy by ensuring that the generated factors optimally represent the document data.

This research is important because it provides insights into how topics related to the new capital, Nusantara, are discussed in Kompas. Kompas online media is one of the Indonesian-language mass media used as a source of news by the Indonesian people (Fahrimal et al., n.d.; Fiorentina et al., 2018). By applying NMF and the Frobenius norm, we can extract dominant themes and trace topic shifts over time in especially about Nusantara Capital City in Kompas. This helps in understanding public and media perceptions of this significant Indonesia national project, identifying issues that are perceived as crucial. Furthermore, this approach can lay the groundwork for further research on NMF-based topic modelling in Indonesian texts, particularly in news analysis related to public policy and national issues.

METHODS

To address the research problem of understanding public and media perceptions of Nusantara, this study adopts topic modeling as its core analytical framework. Specifically, Non-negative Matrix Factorization (NMF) is employed due to its ability to extract coherent and interpretable themes from large textual datasets, such as news articles. This methodology is critical for identifying dominant narratives and shifts in discourse, enabling an in-depth exploration of how Kompas frames issues related to Nusantara. While the Frobenius norm is utilized to optimize the accuracy of the NMF model, technical details regarding its implementation are deferred to the methodology section to maintain focus on the study's broader significance.

By combining NMF's interpretive strengths with the contextual richness of *Kompas* news articles, this research aims to achieve both practical and theoretical outcomes. Practically, it seeks to provide insights into public and media attitudes toward Indonesia’s capital relocation, informing policymakers and stakeholders about key concerns and sentiments. Theoretically, this study contributes to advancing the application of NMF-based topic modeling in analyzing Indonesian-language texts, particularly in domains such as public policy and media studies. Through this dual focus, the research underscores the relevance of its methodologies in uncovering patterns within complex sociopolitical narratives.

Data Collection and Pre-processing; The data used was comprehensive dataset of news articles related to Ibu Kota Nusantara from Kompas.com via web scraping using Python 3. Unwanted characters, symbols, HTML tags and stopwords are thus eliminated from the text so that the data can be used for further processing. Continued by normalization of the data by converting it to lowercase and using stemming or lemmatization to standardize terms (Kartika et al., 2023). Tokenizing is the last step of pre-processing. That is converting each document into a sequence of terms or words that will be analysed in the topic modelling process (Apriani & Gustian, 2019). Based on text mining in March 2024, a collection of news about IKN in Kompas.com was obtained. The number of news collected is 448 news published in the range of January 9th-March 24th 2024. Figure 1 shows a snippet of data or documents in this study. The information collected is the URL of the news, the date of publication, the title, the author, and the content of the news.

Upon the completion of pre-processing, the subsequent phase involves feature extraction. This procedure involved the application of TF-IDF. the Term Frequency-Inverse Document Frequency (TF-IDF) method to represent documents as matrices. Each document is represented as a vector of word frequencies weighted by their

inverse document frequency, resulting in a sparse matrix. The subsequent phase was Non-negative Matrix Factorization (NMF) Model Setup. This phase consist of selecting NMF parameters: n_components=5, random_state=1, init = "nndsvda", beta_loss="frobenius", alpha_W=0.00005, alpha_H=0.00005, l1_ratio=1.

Each parameter used can be explained as follows: n_components=5 is used to specify the number of components or topics to be extracted from the document. In this context, a value of 5 indicates that the model will identify five main topics in the document collection. Research results by (Egger & Yu, 2022a) shows that the choice of the number of components affects the interpretability of the results, where too few components can lose detail, while too many can result in too granular topics. The random_state=1 parameter guarantees the reproducibility of the results by assigning the same initial value to the algorithm, thus allowing the experiment to produce consistent outputs. This was also conveyed in the research by (Vadlapati, n.d.) emphasizing the importance of random state settings in machine learning-based experiments to ensure retestable and comparable results.

The evaluation of topics is conducted through the observation of topic coherence. Coherence scores evaluate the interpretability and semantic relevance of generated topics (Mifrah, 2020b). Increased coherence signifies more significant topic groupings. Evaluation is conducted manually as well. Examine a selection of high-weighted terms within each topic and review sample documents related to these topics to verify their relevance to the Ibu Kota Nusantara theme. A Comparative Analysis is conducted to ensure the model generates relevant topics.

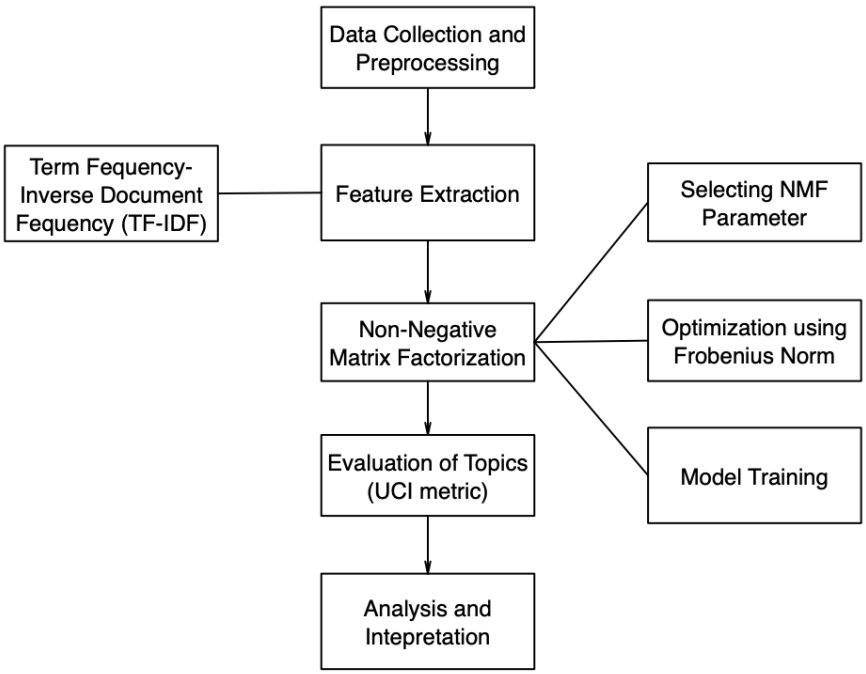

Figure 1. The Research Methodology

FINDINGS AND DISCUSSION

Based on text mining in March 2024, a collection of news about IKN in Kompas.com was obtained. The number of news collected is 448 news published in the range of January 9th-March 24th 2024. Figure 1 shows a snippet of data or documents in this study. The information collected is the URL of the news, the date of publication, the title, the author, and the content of the news.

Mounted at /content/drive

	Unnamed: 0	url	tanggal	judul	penulis	Berita
0	0	https://ikn.kompas.com/read/2024/03/24/0600004...	Kompas.com - 24/03/2024, 06:00 WIB	Dua Investor Malaysia Berencana Bangun Apartem...	Masya Famely Ruhulessin,\nHilda B Alexander	NUSANTARA, KOMPAS.com - Dua investor asal Mala...
1	1	https://ikn.kompas.com/read/2024/03/23/2153274...	Kompas.com - 23/03/2024, 21:53 WIB	Garap 12 Proyek Sekaligus di IKN, PTPP Pastika...	Masya Famely Ruhulessin,\nHilda B Alexander	NUSANTARA, KOMPAS.com - PT PP (Persero) Tbk at...
2	2	https://ikn.kompas.com/read/2024/03/23/1000004...	Kompas.com - 23/03/2024, 10:00 WIB	Jalan Sumbu Kebangsaan Sisi Timur IKN Berlanju...	Masya Famely Ruhulessin,\nHilda B Alexander	NUSANTARA, KOMPAS.com - Pembangunan Jalan Sumb...
3	3	https://ikn.kompas.com/read/2024/03/23/0900004...	Kompas.com - 23/03/2024, 09:00 WIB	Volume Lalin Tol Balsam Bakal Naik Saat Mudik,...	Masya Famely Ruhulessin,\nHilda B Alexander	BALIKPAPAN, KOMPAS.com - Jasamarga Nusantara T...
4	4	https://ikn.kompas.com/read/2024/03/23/0800009...	Kompas.com - 23/03/2024, 08:00 WIB	Masih Ada Waktu Sosialisasi, Tak Ada Alasan Ol...	Masya Famely Ruhulessin,\nHilda B Alexander	NUSANTARA, KOMPAS.com - Ombudsman RI menilai O...
...	...	...	...	...	...	...
143	143	https://ikn.kompas.com/read/2024/01/24/0917438...	Kompas.com - 24/01/2024, 09:17 WIB	Karena IKN, Okupansi Mal Terbesar di Kalimanta...	Masya Famely Ruhulessin,\nHilda B Alexander	SAMARINDA, KOMPAS.com - Relokasi ibu kota nega...

Figure 2. News Dataset Snippet from Kompas

The subsequent step involves pre-processing, initiated by encompasses the normalization of words. This step ensures uniform representation of all characters in the text. Text processing requires the word normalization process to ensure that the text is converted into a uniform format, facilitating readability and further processing. Selanjutnya, adalah menghilangkan karakter special seperti `[-()\[\]\"/@;%*:<>{}`+=~|.!?,&].` Removing such characters from text is done in the process of cleaning up the text. The goal is to improve data consistency and make text easier for algorithms to process. The final stage of pre-pocessing was to omit certain words on the document. The list of omitted words are:`'juga','dan','dengan','untuk','pada','akan','dari','itu','yang','content','continue','ADVERTISEMENTSCROLL TO CONTINUE WITH CONTENT'`. In this study, the list of words is called stopwords. Eliminating stopwords is a common step in text preprocessing. Stopwords are often most words in a text document. If not removed, they will still be a feature in vector representations, despite having a low weight in TF-IDF. By removing stopwords, it can reduce the data dimension, thereby speeding up the computing process. Figure 3 shows the result of pre-processing.

	Berita
0	NUSANTARA Dua investor asal Malaysia IJM Corpo...
1	NUSANTARA PT PP Persero Tbk PTPP memastikan du...
2	NUSANTARA Pembangunan Jalan Sumbu Kebangsaan S...
3	BALIKPAPAN Jasamarga Nusantara Tollroad Region...
4	NUSANTARA Ombudsman RI menilai Otorita Ibu Kot...
...	...
143	SAMARINDA Relokasi ibu kota IKN Nusantara Kali...
144	BALIKPAPAN Murah senyum ramah aktif berbicara ...
145	BALIKPAPAN Penumpang memanfaatkan layanan anta...
146	JAKARTA Pemerintah Lembaga Manajemen Aset Nega...
147	JAKARTA Kementerian Pekerjaan Umum Perumahan R...

448 rows × 1 columns

Figure 3. Preprocessing Results Snapshot

```
Loading dataset...
done in 0.000s.
Extracting tf-idf features for NMF...
done in 0.103s.
  (0, 584)      0.061320590866633315
  (0, 207)      0.06083353944236541
  (0, 295)      0.11747986865894888
  (0, 42)       0.059081901590353945
  (0, 464)      0.45169458446828553
  (0, 234)      0.16273973467252198
  (0, 138)      0.2308489309845647
  (0, 927)      0.057073432206800966
  (0, 494)      0.03538961227933092
  (0, 265)      0.12846876744178437
  (0, 34)       0.05188543725450244
  (0, 269)      0.03135289019470335
  (0, 553)      0.054246578224174
  (0, 425)      0.05325510609871522
  (0, 115)      0.15981745941899062
  (0, 279)      0.11230552236138204
  (0, 832)      0.03989707399011127
  (0, 707)      0.19761894736522176
  (0, 975)      0.21674877370657972
  (0, 98)       0.12119200050449824
  (0, 484)      0.10414443030915435
  (0, 294)      0.18048834538755779
  (0, 830)      0.1651426561192403
  (0, 642)      0.22584729223414277
  (0, 937)      0.08095167186691446
```

Figure 4. TF-IDF Results Snapshot

The feature extraction process is carried out by applying TF-IDF. Term Frequency-Inverse Document Frequency (TF-IDF) is a technique used to assess the importance of a word in a document relative to other collections of documents (corpus). This method is used to represent text as numerical data that can be processed by NMF model

(Egger & Yu, 2022b). Term Frequency (TF) is a measure of how often a word appears in a document. Inverse Document Frequency (IDF) is a measure of how unique or rare a word appears throughout a collection of documents (Kartika et al., 2023). Figure 4 shows the results of the TF-IDF calculation on the document. At this stage, the document is ready for topic modeling with Non-Negative Matrix Factorization (NMF). Initiated from setting Non parameters and using Frobenius Norm and then training the model.

Non-Negative Matrix Factorization (NMF) is a matrix decomposition-based dimension reduction technique used to factor matrix A into two matrices W and H (Woo et al., 2020) where all values in this matrix are non-negative (Kang et al., 2023). NMF is widely used in machine learning, text mining, and computer vision for big data analysis with simpler interpretations (Egger & Yu, 2022b). The use of the Frobenius Norm is to measure reconstruction errors. Frobenius Norm is a metric that is often used to measure errors in various applications such as matrix approximation, linear equation solution, and optimization (Peng, 2015). In the context of Non-Negative Matrix Factorization (NMF), Frobenius Norm is used to evaluate the difference between the original matrix and the factorization result. The results of topics using NMF are shown in Figure 5.

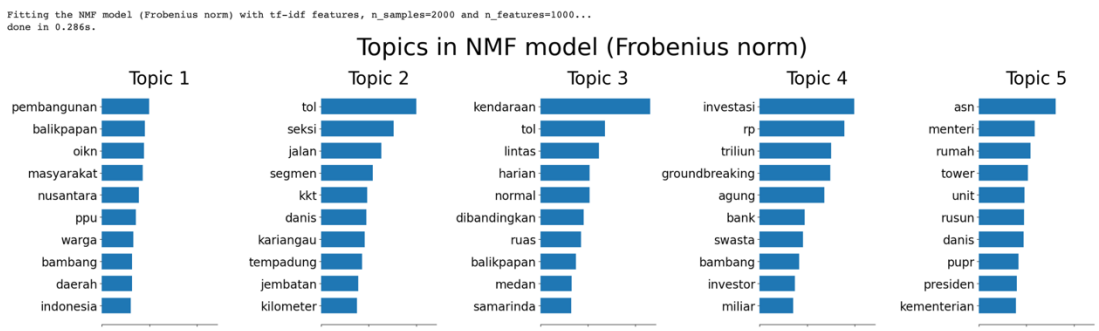

Figure 5. The distribution of topics in Kompas documents related to IKN news.

In Figure 5, There are five topics derived from terms linked to the issue. It can be seen that each topic formed has constituent words that are relevant to each other. It can be seen that the Kompas.com discussion about the IKN in the January-March 2024 range is about the preparation of the IKN community, infrastructure development, traffic conditions in the IKN, groudbreaking in the IKN, and accommodation for ASN. In detail, the interpretation of the discussion on each topic is conveyed through Table 1.

Table 1. Human Intrepretation of each topics

Topics	Kata-kata dalam topik	Human intepretation
Topic 1	*Pembangunan, Balikpapan, oikn, masyarakat, nusantara, ppu, warga, bambang, daerah Indonesia*	Topic regarding the preparation of the community/residents in IKN
Topic 2	*Tol, seksi, jalan, segmen, kkt, danis, karingau, tempadung, jembatan, kilometer*	Topic regarding information on toll road development updates in IKN
Topic 3	*Kendaraan, tol, lintas, harian, normal, dibandingkan, ruas, Balikpapan, medan, samarinda*	Topic regarding the condition of the IKN buffer section, namely the Balikpapan-Samarinda (Balsam) toll road

Topics	Kata-kata dalam topik	Human intepretation
Topic 4	*Investasi, rp, triliun, grounbreaking, agung, bank, swasta, Bambang, investor, miliar*	Topic About Groundbreaking in IKN
Topic 5	*Asn, Menteri, rumah, tower, unit, rusun, danis, pupr, presiden, kementerian*	Topic regarding ASN flats and office houses in IKN

To better view the relevance between the topic and the document, a manual check is carried out on the related news. Table 2 shows the news related to each topic as well as the main point if the news.

Table 1. Topic Evaluation of Related Documents

Topics	News	The main point of the news
Topic 1 Regarding the preparation of the community/residents in IKN	Related News example: *"Wujudkan Transportasi Ramah Lingkungan di IKN, Bluebird Koordinasi dengan Otorita"* News Link: https://ikn.kompas.com/read/2024/02/15/183324987/wujudkan-transportasi-ramah-lingkungan-di-ikn-bluebird-koordinasi-dengan-otorita.	PT Blue Bird Tbk supports the development of environmentally friendly transportation in the Nusantara Capital City (IKN) in coordination with OIKN to provide environmentally friendly EV-based services. This effort aims to increase community mobility in the Penajam Paser Utara (PPU) area as part of the development of IKN infrastructure.
Topic 2 Regarding information on toll road development updates in IKN	Related News example: *"Tembus 90 Persen, Tiga Ruas Tol IKN Siap Sambut Jokowi"* News Link: https://ikn.kompas.com/read/2024/07/06/090000987/tembus-90-persen-tiga-ruas-tol-ikn-siap-sambut-jokowi-.	The progress of the construction of three IKN toll roads, namely Section 3A Karang Joang-KKT Kariangau, Section 3B KKT Kariangau-Sp. Tempadung, and Section 5A Sp. Tempadung-Balang Island Bridge, have reached 90 percent. Danis Hidayat Sumadilaga, Chairman of the IKN Infrastructure Development Implementation Task Force, ensured that this toll road would be ready to be used in one direction with three lanes ahead August 17, 2024.
Topic 3 regarding the condition of the IKN buffer section, namely the Balikpapan-Samarinda (Balsam) toll road	Related News example: *"Naik 100 Persen, Kendaraan Lintasi Tol Balsam di H+3 Lebaran"* News Link: https://ikn.kompas.com/read/2024/04/14/151454187/naik-100-persen-kendaraan-lintasi-tol-balsam-di-h3-lebaran.	The volume of vehicles on the Balikpapan-Samarinda (Balsam) toll road reached 21,649 units on D+3 Eid 1445 H, an increase of 104.2 percent compared to normal daily traffic of 10,601 vehicles. Similar increases also occurred on several other toll roads outside Java during the long holiday period. Jasamarga Nusantara Tollroad appeals to motorists to ensure that their vehicles are in prime condition and take advantage of rest area facilities on each toll road.
Topic 4 About Groundbreaking in IKN	Related News example: *"Kamis Ini, Jokowi Dijadwalkan*	Jokowi is scheduled to lead the groundbreaking of three banking projects in IKN, including Bank Mandiri, BNI, and BRI offices, as well as the signing of

	"Groundbreaking" Tiga Proyek Bank" News Link: https://ikn.kompas.com/read/2024/02/29/060000187/kamis-ini-jokowi-dijadwalkan-groundbreaking-tiga-proyek-bank.	cooperation between OJK and OIKN. This project is part of phase V, with an estimated investment value of close to Rp 50 trillion. OIKN Deputy, Agung Wicaksono, emphasized that until January 2024, the investment entering the IKN has reached Rp 47.5 trillion, with additional value from this latest project.
Topic 5 Regarding ASN flats and office houses in IKN	Related News example: *"Serba Canggih, Rusun ASN di IKN Gunakan Teknologi Smarthome"* News Link: https://ikn.kompas.com/read/2024/08/31/161400687/serba-canggih-rusun-asn-di-ikn-gunakan-teknologi-smarthome.	ASN flats in the IKN will be equipped with smart home technology, such as digital locks and electricity and gas control through smart devices, according to Danis H. Sumadilaga. A total of 47 towers with 60 units per tower are targeted to be completed in December 2024, while 21 towers are expected to be completed in September-October. The Minister of PUPR, Basuki Hadimuljono, explained that ASN with a family will get one flat unit, while single ASN can share a unit.

Table 1 shows that all topics generated by the NMF model have good relevance to the IKN and the words that make up the topic are relevant to each other in each topic. To further evaluate, topic coherence metrics are used.

Topic coherence is employed to assess the model's capability in generating topics. Topic coherence metrics estimate the degree of similarity across latent topics identified by a topic model from a human-like perspective (Newman et al., 2011), aiming to ascertain a high level of semantic coherence in topic models. Topic coherence evaluates the scope of a topic by calculating word co-occurrences and mutual information, which can indicate how individuals perceive the issue (Blair et al., 2020; Campagnolo et al., 2022). This study used *c_uci* coherence score. The coherence score is derived from sliding windows and the pointwise mutual information of all word pairs, utilizing the top *N* words by frequency (Zvornicanin, 2024). To comprehend the concept of *c_uci*, it can be observed in Campagnolo et al. (2022). A *c_uci* score ranges from 0 to 1, where higher values indicate more coherent and human-interpretable topics (Stevens et al., 2012). In this study, the result of the *c_uci* is 0.9912595238288311. This number can be expressed as **High Coherence**. A *c_uci* score of 0.991 suggests that the topics are reasonably coherent. The top words within each topic are likely semantically related, making the topics easier for humans to interpret.

The next evaluation was using wordcloud. A wordcloud is a visual representation of the frequency or relevance of a word in a text. Words that appear more often in text are given larger or more prominent letter sizes. Wordcloud in this study is used to provide a glimpse of the main themes in the text. Wordcloud does not show relationships between words (for example, co-occurrence or context). Words with slightly higher frequencies can dominate the display, although not significant. The results of the wordcloud of corpus that has been cleaned (has gone through preprocessing) are shown in Figure 6.

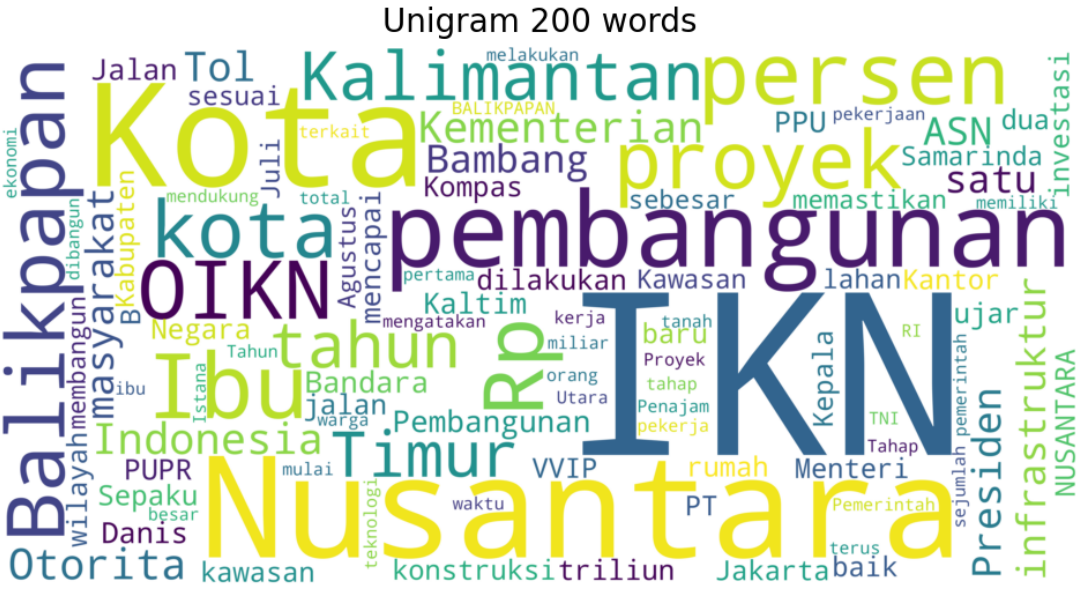

Figure 6. The World Cloud of terms of IKN news in Kompas.com

In Figure 6, the large words are: IKN, archipelago, development, OIKN, Balikpapan, Kalimantan, project, percent, city, Rp, and so on. Figure 6 is the result of Wordcloud creating visualizations based on the frequency of words in the text. Words that appear more often are given a larger visual representation. When compared with the topic results from NMF, It can be seen that the same words appear. This happens because NMF reduces the dimension of text data by decomposition of the TF-IDF matrix so that each topic produced is a linear combination of words with a certain weight. In NMF, the words with the highest weight for each topic are considered the dominant words. NMF uses a matrix to show the words that contribute the most to the topic. These words are often high-frequency words in related documents, so they match the Wordcloud results. Wordcloud can help to understand the content or theme of a topic quickly without having to check word lists or numerical weights. The consistency between Wordcloud and NMF results occurs because these two methods prioritize the words that matter in the dataset, albeit with different approaches. Wordcloud helps visualize dominant words, while NMF provides a more in-depth structural framework for understanding the relationships of words in a topic.

CONCLUSION

This study yields the following conclusions:

Initially, the application of NMF resulted in the formation of five distinct topics. Each topic is composed of constituent words that are interrelated. The news from Kompas.com concerning the IKN during the January-March 2024 period focuses on the topic of community and resident preparations in IKN. Discussion concerning the latest developments in toll road construction in IKN. Discussion concerning the status of the IKN buffer section, specifically the Balikpapan-Samarinda (Balsam) toll road. Discussion on the groundbreaking in IKN and the subject of ASN flats and office houses in IKN. The topics produced by the NMF model demonstrate strong relevance to the IKN, with the words within each topic showing significant interrelatedness.

Manual searching reveals that these topics are indeed being discussed in the news regarding IKN in the Kompas media.

Secondly, topic coherence is utilized to evaluate the model's effectiveness in generating topics. This research employed the c_uci coherence score. The outcome of the coefficient of variation in this study is 0.9912595238288311. This number can be articulated as High Coherence. A c_uci score of 0.991 indicates a satisfactory level of coherence among the topics. The leading terms associated with each topic are probably semantically connected, thereby facilitating human interpretation of the topics.

Third, the alignment between the Wordcloud and NMF results can be attributed to the fact that both methods emphasize the significant words within the dataset, although they employ distinct methodologies. Wordcloud serves to illustrate the prevalent words, whereas NMF offers a comprehensive structural framework for analyzing the interconnections of words within a topic.

REFERENCES

Alfajri, A., Richasdy, D., & Bijaksana, M. A. (2022). Topic Modelling Using Non-Negative Matrix Factorization (NMF) for Telkom University Entry Selection from Instagram Comments. *Journal of Computer System and Informatics (JoSYC)*, *3*(4). https://doi.org/10.47065/josyc.v3i4.2212

Ambarwati Puspitasari, D., Karlina, Y., & Mulyo, B. M. (2023). Tracing Trends and Patterns of IKN Words in Media and Twitter: A Linguistic Corpus Study. *Jurnal Penelitian Pendidikan Bahasa Dan Sastra Indonesia*, *9*, 239–257. https://doi.org/10.22202/jg.v9i2.6947

Amelia, R. (2023). Sentiment Analysis of Government Policy in Relocating the Republic of Indonesia's Capital City. *Journal of Public Administration and Government*, *5*(2). https://jurnal.fisip.untad.ac.id/index.php/JPAG

Apriani, R., & Gustian, D. (2019). Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*, *6*(1), 54–62.

Astuti, R. H., Muljono, M., & Sutriawan, S. (2024). Indonesian News Text Summarization Using MBART Algorithm. *Scientific Journal of Informatics*, *11*(1), 155–164. https://doi.org/10.15294/sji.v11i1.49224

Athukorala, S., & Mohotti, W. (2022). An effective short-text topic modelling with neighbourhood assistance-driven NMF in Twitter. *Social Network Analysis and Mining*, *12*(1). https://doi.org/10.1007/s13278-022-00898-5

Blair, S. J., Bi, Y., & Mulvenna, M. D. (2020). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, *50*(1). https://doi.org/10.1007/s10489-019-01438-z

Campagnolo, J. M., Duarte, D., & Dal Bianco, G. (2022). Topic Coherence Metrics: How Sensitive Are They? *Journal of Information and Data Management*, *13*(4). https://doi.org/10.5753/jidm.2022.2181

Clerwall, C. (2014). Enter the Robot Journalist. *Journalism Practice*, *8*(5), 519–531. https://doi.org/10.1080/17512786.2014.883116

Dhamayanti, L. P., Mani, L., & Dhamayanti, M. (2024). Online News: Media Framing On Indonesia's Capital City Relocation Policy. *Dirasat: Human and Social Sciences*, *51*(6). https://doi.org/10.35516/hum.v51i6.3

Egger, R., & Yu, J. (2022a). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*. https://doi.org/10.3389/fsoc.2022.886498

Egger, R., & Yu, J. (2022b). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*. https://doi.org/10.3389/fsoc.2022.886498

Fahrimal, Y., Husna, A., Islami, F., Studi Ilmu Komunikasi, P., Teuku Umar Jl Alue Peuyareng, U., Utu, K., & Barat, A. (n.d.). *MEDIA DAN PANDEMI: FRAME TENTANG PANDEMI COVID-19 DALAM MEDIA ONLINE DI INDONESIA (Studi Pada Portal Berita Kompas.com dan Detik.com) MEDIA AND PANDEMIC: FRAME OF COVID-19 PANDEMIC IN INDONESIAN ONLINE MEDIA (Study on Kompas.com and Detik.com News Websites)*.

Fiorentina, R., Mayasari, M., & Hariyanto, F. (2018). Analisis Framing Pemberitaan "Reuni Akbar 212" (Analisis Framing Model Robert N Entman Media Online kompas.com dengan republika.co.id Edisi 26 November 2017 – 9 Desember 2017). *Jurnal Politikom Indonesiana*, *3*(2). https://doi.org/10.35706/jpi.v3i2.1657

Kang, M., Zhao, J., & Han, Z. (2023). Robustness of NMF Algorithms Under Different Noises. *EAI Endorsed Transactions on Internet of Things*, *9*(1). https://doi.org/10.4108/eetiot.v9i1.3271

Kartika, L. G. S., Utama, P. K. L., Budiastawa, I. D. G., & Rinartha, K. (2023). Comparison of the Sentiment Analysis Model's Code Complexity and Processing Time. *Sinkron*, *8*(1). https://doi.org/10.33395/sinkron.v8i1.11894

Lotto, M., Hussain, I. Z., Kaur, J., Butt, Z. A., Cruvinel, T., & Morita, P. P. (2023). Analysis of Fluoride-Free Content on Twitter: Topic Modeling Study. *Journal of Medical Internet Research*, *25*. https://doi.org/10.2196/44586

Mifrah, S. (2020a). Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using COVID'19 Corpus. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4). https://doi.org/10.30534/ijatcse/2020/231942020

Mifrah, S. (2020b). Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using COVID'19 Corpus. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 5756–5761. https://doi.org/10.30534/ijatcse/2020/231942020

Newman, D., Bonilla, E. V, & Buntine, W. (2011). Improving Topic Coherence with Regularized Topic Models. Improving Topic Coherence with Regularized Topic Models. *Advances in Neural Information Processing Systems 24 (NIPS 2011)*. https://www.researchgate.net/publication/260639294

Nurjayanto, B. W., & Baizal, Z. K. A. (2024). NEWS RECOMMENDER SYSTEM USING HYBRID CONTENT-BASED FILTERING AND COLLABORATIVE FILTERING. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *9*(1), 26–33. https://doi.org/10.29100/jipi.v9i1.4256

Pan, S., Chen, J., Ganesan, A., & Brantley, K. (2015). LDAExplore: Visualizing Topic Models Generated Using Latent Dirichlet Allocation. *ArXiv Preprint ArXiv:1507.06593*. https://doi.org/https://doi.org/10.48550/arXiv.1507.06593

Peng, Y. (2015). Inequalities for the Frobenius norm. *Journal of Mathematical Inequalities*, *9*(2), 493–498. https://doi.org/10.7153/jmi-09-43

Samuels, A., & Mcgonical, J. (2019). *Sentiment Analysis of News Articles: A Lexicon based Approach*. http://mlg.ucd.ie/datasets/bbc.html.

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). *Exploring Topic Coherence over many models and many topics*. Association for Computational Linguistics. http://mallet.cs.umass.edu/

Steyvers, M., & Griffiths, T. (2007). Probabilistic Topic Models. In *Handbook of Latent Semantic Analysis*. Laurence Erlbaum.

Sutherland, I., Sim, Y., Lee, S. K., Byun, J., & Kiatkawsin, K. (2020). Topic modeling of online accommodation reviews via latent dirichlet allocation. *Sustainability (Switzerland)*, *12*(5). https://doi.org/10.3390/su12051821

Vadlapati, P. (n.d.). Does Seed Matter? Investigating the Effect of Random Seeds on LLM Accuracy. *International Journal on Science and Technology (IJSAT) IJSAT23031126*, *14*(3). https://doi.org/10.5281/zenodo.14288248

Woo, H. S., Kim, J. H., Kim, J. M., & Lee, W. G. (2020). Exploring the AI topic composition of K-12 using NMF-based topic modeling. *International Journal on Advanced Science, Engineering and Information Technology*, *10*(4). https://doi.org/10.18517/ijaseit.10.4.12787

Zvornicanin, E. (2024, March). *When Coherence Score Is Good or Bad in Topic Modeling?* Https://Www.Baeldung.Com/Cs/Topic-Modeling-Coherence-Score.