TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 22, No. 4, August 2024, pp. 784~798
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v22i4.26063

Unlocking insights from Ministry of Marine Affairs and Fisheries annual reports using LDA: a deep dive into SDG 14

Ahmad Marzuqi[1], Rezzy Eko Caraka[1,2,3,4,6], Prana Ugiana Gio[5], Rung Ching Chen[6], Maengseok Noh[7], Bens Pardamean[8,9]

[1]Faculty of Economics and Business, Campus UI Depok, Universitas Indonesia, Depok, Indonesia
[2]School of Economics and Business, Telkom University, Bandung, Indonesia
[3]Research Center for Data and Information Sciences, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Bandung, Indonesia
[4]Department of Mathematical Sciences, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[5]Department of Mathematics, Universitas Sumatera Utara, Medan, Indonesia
[6]Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan
[7]Data and Information Sciences College of Information Technology and Convergence, Pukyong National University, Busan, Republic of Korea
[8]Department of Computer Science, BINUS Graduate Program-Master of Computer Science Program, Bina Nusantara University, Jakarta, Indonesia
[9]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Feb 8, 2024
Revised Feb 28, 2024
Accepted Mar 29, 2024

Keywords:

Annual reports
Fisheries
Latent Dirichlet allocation
Marine resources
Ministry of Marine Affairs and Fisheries
Strategic planning
Topic modeling

ABSTRACT

Annual reports serve as vital instruments for government ministries and agencies, enabling transparency and accountability in managing state budgets (APBN) and activities, thereby fulfilling a crucial role in public accountability, particularly in the context of sustainable development goal (SDG) 14. However, due to their extensive nature, it becomes imperative to conduct topic modeling analysis to discern trends and topics within these reports. In this study, latent Dirichlet allocation (LDA), a prominent topic modeling technique, is employed to analyze the annual reports of the Ministry of Marine Affairs and Fisheries (KKP) Indonesia from 2015 to 2022. Utilizing the coherence score as an evaluation metric, we assess the quality of topic models across each report year. Our findings underscore the consistent emphasis on fisheries and marine-related initiatives, emphasizing their relevance to SDG 14 and Indonesia's maritime landscape. Ultimately, this study offers valuable insights to inform strategic planning and decision-making processes within the KKP, contributing to the advancement of SDG 14 and promoting sustainable development in Indonesia's fisheries and marine sectors.

Corresponding Author:

Rezzy Eko Caraka
Research Center for Data and Information Sciences, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN)
Bandung, Indonesia
Email: rezzy.eko.caraka@brin.go.id or rezzyekocaraka@telkomuniversity.ac.id

1. INTRODUCTION

The evolution of the annual report within the corporate design landscape has transformed it into a sophisticated product. Rather than merely serving as a repository of information, its primary function now extends to proactively shaping visibility and conveying specific meanings [1]-[3]. In essence, the modern annual report aims to create a narrative that goes beyond the surface level of *"what's in the document"*. It strategically employs design elements, such as layout, typography, imagery, and branding, to communicate

key messages, values, and achievements effectively. Through thoughtful design choices and strategic storytelling, organizations can leverage their annual reports as powerful tools for building trust, engaging stakeholders, and shaping perceptions in the competitive business environment [4], [5]. The annual report encapsulates a comprehensive array of information, including financial data, strategic accomplishments, forward-looking projections, corporate governance practices, and the social and environmental ramifications attributable to the organization's operations [6]. In Indonesia, adherence to regulatory requirements mandates that issuers or public companies furnish their annual reports to the Financial Services Authority (OJK) within the stipulated timeframe, typically by the conclusion of the fourth month subsequent to the closure of the financial year. This regulatory framework underscores the significance placed on transparency, accountability, and timely dissemination of information within the Indonesian business landscape, facilitating informed decision-making processes among stakeholders and fostering trust, and confidence in the financial markets [7]. While the primary objective of an annual report is indeed to offer stakeholders, including shareholders, employees, government entities, and the general public, a holistic view of a company's performance and standing throughout a specified period, it's true that many stakeholders may not fully capitalize on its potential for deeper analysis [8]. Despite being rich repositories of valuable information, annual reports often face underutilization due to various factors. One reason for this underutilization may be the sheer volume and complexity of information contained within annual reports. Stakeholders may find it challenging to navigate through extensive financial statements, strategic narratives, and governance disclosures to extract actionable insights. Furthermore, stakeholders may lack the requisite financial literacy or expertise to interpret the nuanced details presented in annual reports effectively. Without a thorough understanding of financial metrics, industry dynamics, and corporate governance principles, stakeholders may struggle to derive meaningful conclusions from the information provided. Additionally, stakeholders may perceive annual reports as dense, formal documents primarily designed for regulatory compliance rather than as strategic communication tools. As a result, they may overlook the potential insights and opportunities for analysis that annual reports can offer [9].

To address these challenges and enhance stakeholder engagement with annual reports, companies can adopt several strategies. These include simplifying the presentation of complex information, providing clear explanations and contextualizations of key metrics, and leveraging technology to enhance accessibility and interactivity. Moreover, companies can proactively engage stakeholders through supplementary communications, such as investor presentations, executive summaries, and interactive online platforms, to highlight key highlights and facilitate deeper understanding and analysis of annual report content. By enhancing the usability and relevance of annual reports, companies can empower stakeholders to make informed decisions and foster greater transparency and trust in corporate reporting practices [1], [9]. Analyzing such high-volume reports has traditionally involved time-consuming reading and interpretation [10]-[12]. In a typical company annual report, textual narratives represent the bulk of disclosures with an average of 80% of the annual report, compared to the remainder consisting of numbers and quantitative data representations [13]. Indeed, technological advancements, particularly in the realm of machine learning and natural language processing, have revolutionized the analysis of annual reports, making it both rapid and objective. Topic modeling techniques, such as latent Dirichlet allocation (LDA), enable the swift extraction of key themes and insights from voluminous documents, facilitating a deeper understanding of the underlying content. These automated approaches offer several advantages over manual methods, including speed, scalability, and objectivity. By swiftly identifying hidden patterns and themes within annual reports, topic modeling techniques provide stakeholders with valuable insights into a company's performance, strategic priorities, and prospects.

The novelty of this paper lies in its application of text mining techniques to analyze the content of the annual reports issued by the Ministry of Maritime Affairs and Fisheries (KKP). While annual reports are traditionally viewed as static documents containing predefined sections such as financial statements and management discussions, this study harnesses the power of text mining to unearth hidden insights and identify emergent themes within the reports. By employing text mining methodologies, the study transcends conventional approaches to annual report analysis, which often rely on manual review and interpretatio [14]. By using the LDA method to find the main topics [15].

Related research [16], [17] did topic modeling research use LDA. Next [17], using metadata from all peer-reviewed articles published in International Journal of Lifelong Education (1982–2021). Also, [16] conducted topic modeling research using a dataset in the form of a collection of scientific publications from the University of Nairobi, Kenya. The data consists of abstracts extracted from 10.000 publication documents.

Another researchs [18]-[20], conducted research related to the development of the LDA method. In research [18], the methods used in the research are the development of an interval semi-supervised modification of the LDA model, referred to as ISLDA, and the creation of a new topic quality metric called tf-idf coherence. Whereas in the research [21] the researcher developed the author-topic model, which is a generative probabilistic model designed to explore the relationships between authors, documents, topics, and words. Indeed, the author-topic model is a valuable probabilistic model that enables the exploration of relationships

between authors, documents, topics, and words within a corpus of text data. This model extends the traditional topic modeling framework, such as LDA, by incorporating authorship information into the modeling process.

Our paper is structured into several distinct sections to effectively convey its research findings. Section 1 serves as the introduction, offering a comprehensive overview of the research problem, its significance, and the objectives pursued. Following this, section 2 delves into the system overview and methods employed, particularly focusing on the application of LDA for topic modeling. It elucidates the theoretical underpinnings of LDA and outlines its implementation within the research framework. In section 3, attention shifts towards the dataset utilized in the study and the preprocessing techniques applied to ensure data quality and relevance. This section details the source, size, and pertinent characteristics of the dataset, along with the steps taken to clean and prepare the data for analysis, such as normalization and tokenization. Subsequently, section 4 presents the results of the analysis, accompanied by a thorough discussion of the findings. Here, the paper elucidates the main topics extracted from the dataset and interprets their significance within the context of the research objectives. Finally, section 5 encapsulates the conclusion, summarizing the key insights gleaned from the study and highlighting avenues for future research. Through this structured approach, the paper systematically elucidates its method, findings, and implications, thereby contributing valuable knowledge to the field of topic modeling and text analysis.

2. TOPIC MODELLING

Topic modeling is a text analysis technique that aims to identify and extract the main topics or hidden semantic structures in a document [22]. Pattern discovery often reflects the underlying topics that come together to form documents, such as hierarchical probabilistic models that are easily generalized to other types of data. Topic modeling has been used to analyze things other than words such as images, biological data, and survey information and data [23]. The goal is to present hidden concepts, salient features or latent variables from the data, depending on the application context, to be identified efficiently [24].

Currently, topic modeling presents a number of varied methods for text analysis. Four methods that topic modeling often relies on are latent semantic analysis (LSA), probabilistic latent semantic analysis (PLSA), LDA, correlated topic model (CTM) [25]. Based on research [26], LDA has advantages over PLSA for classification tasks that require a latent semantic representation with good granularity that can be generalized from training data to test data. LDA introduces an element of probability into the model by assuming that a Dirichlet distribution underlies the distribution of topics in documents. Therefore, this research uses LDA as a topic modeling method. LDA is a generative probabilistic model of a corpus [27]. The basic idea is that documents are described as a random mixture of latent topics, with a unique distribution of words for each topic. LDA assesses documents as a complex collaboration between topics, creating a framework that allows for a deeper understanding of the semantic representation in the corpus.

LDA is an effective method in analyzing large-scale documents or texts. The advantage of LDA lies in its ability to summarize, connect, and process large volumes of data. This model can generate a list of topics with associated weights for each document [28] as such, LDA is not only an effective analysis tool, but also an invaluable method for exploring information from texts with a high degree of complexity. The likelihood of observing the entire corpus of documents in LDA, denoted as $P(corpus \mid \alpha, \beta)$ is expressed as the product over all documents in the corpus. Mathematically, it is represented as (1):

$$P(corpus \mid \alpha, \beta) = \prod d \int p(\theta_d|\alpha) \left(\prod n \prod z \, p \left(z_{d,n}|\theta_d\right) p\left(w_{d,n}|\phi_z\right)\right) d \, \theta_d \quad (1)$$

Here, $p(\theta_d|\alpha)$ represents the Dirichlet distribution over the topic proportions for document d, $p\left(z_{d,n}|\theta_d\right)$ represents the probability of topic z given the document's topic proportions θ_d. In line with this, $p\left(w_{d,n}|\phi_z\right)$ represents the probability of word $w_{d,n}$ given the word distribution ϕ_z for topic z. This likelihood function is fundamental in estimating the parameters of the LDA model from observed data [19], [29]. Techniques such as variational inference or Gibbs sampling are commonly employed to approximate this likelihood and infer the posterior distributions over the latent variables, enabling the discovery of latent topics within the corpus.

The likelihood involves integrating over the latent variable θ_d which captures the distribution of topics within document d and summing over all possible topics z for each word in the document. This integral and summation account for the uncertainty associated with both the topic proportions and the topic assignments for each word, yielding the overall likelihood of observing the entire corpus given the model parameters α and β. Word cloud is a visual representation technique of the frequency of words from a dataset. The more frequently the term appears in the text being analyzed, the larger the word appears in the resulting image [30]. Word clouds are increasingly being used as a simple tool to identify the focus of written material. Word clouds are used to provide an intuitive and visually appealing overview of a text by depicting the words that appear most

frequently in it [31]. These summaries are helpful for learning the number and types of topics present in a text. Usually, this statistical overview is achieved by positively correlating the font size of the tags depicted with the word frequency. Let $f(w)$ represent the frequency of word w in the corpus. Let $s(w)$ represent the font size of word w in the word cloud. Then, a basic equation for determining the font size of each word in the word cloud could be:

$$s(w) = k \times f(w) \tag{2}$$

k is a scaling factor to adjust the font sizes based on the range of word frequencies and the desired size of the word cloud. This equation simply scales the font size of each word proportionally to its frequency. However, in practice we considerations such as logarithmic scaling, normalization, and adjusting for visual aesthetics may be incorporated into the equation to improve the readability and visual appeal of the word cloud. Logarithmic scaling helps to reduce the impact of extremely high word frequencies and provides a smoother distribution of font sizes. We can use the logarithm of the word frequency as the basis for font sizing:

$$s(w) = k \times \log\big(f(w)\big) \tag{3}$$

Normalization ensures that the font sizes are scaled appropriately relative to the maximum word frequency in the corpus. This prevents very high-frequency words from dominating the word cloud. One common normalization technique is to divide each word frequency by the maximum frequency:

$$Norm_{freq}(w) = \frac{f(w)}{Maximum_{freq}} \tag{4}$$

Then, the normalized font size equation becomes:

$$s(w) = k \times \log\big(norm_{freq}(w) + 1\big) \tag{5}$$

By incorporating logarithmic scaling and normalization, the word cloud equation provides a balanced distribution of font sizes that emphasizes both high-frequency words and less frequent terms while ensuring readability and visual appeal. The coherence score serves as a crucial metric in assessing the quality of topics derived from topic modeling algorithms like LDA. While there isn't a single mathematical equation to compute coherence score, its conceptual framework revolves around measuring the semantic similarity between the top words within each topic. Each topic T consists of a set of top words, and the coherence score $Coherence\ (T)$ is calculated as the average pairwise similarity between these words. This computation involves summing up the pairwise similarities for all combinations of top words within each topic and then averaging these values across all topics [32]. The resulting coherence score offers a quantitative assessment of the coherence and interpretability of the topics. Higher coherence scores signify that the topics encapsulate words with closer semantic relationships, indicating a more meaningful representation of themes within the corpus. As such, coherence score serves as a valuable tool in guiding the refinement and evaluation of topic models, facilitating the extraction of meaningful insights from textual data.

Let T represent the set of all topics generated by the LDA model. Let $TopWords(t)$ represent the top words in topic t. Let $Sim\ (w_i, w_j)$ represent a measure of semantic similarity between words (w_i, w_j). Let $Coherence\ (T)$ represent the coherence score for the set of topics T.

$$Coherence\ (T) = \frac{1}{|T|}\sum_{t\in T}\frac{2}{|TopWords(t)|\ (|TopWords(t)|-1)}\sum_{i<j} Sim\ (w_i, w_j) \tag{6}$$

Determining the optimal coherence score in topic modeling is a nuanced endeavor, contingent upon various factors including dataset complexity and research objectives. While a higher coherence score typically indicates more coherent and interpretable topics, there isn't a universally accepted threshold for what constitutes an ideal score. Instead, researchers often compare coherence scores across different models or parameter settings to identify the configuration that best aligns with their analytical goals. In practical terms, coherence scores serve as invaluable guides during model refinement, allowing researchers to fine-tune parameters and track improvements in topic quality. While coherence scores above zero are generally desirable, they should be interpreted alongside other factors such as topic interpretability and relevance to the research context. Ultimately, the determination of the "best" coherence score is context-dependent and requires a comprehensive understanding of the specific dataset and research objectives [6], [8].

3. DATASET

In this study, our topic modeling process system was developed using the annual reports of the KKP spanning from 2015 to 2022. LDA was employed as the topic modeling method. Data preprocessing and topic modeling were conducted using the Python programming language within the Jupyter Notebook software environment. Prior to analysis, the dataset underwent preprocessing to enhance the effectiveness and efficiency of the mining process. The study utilized secondary data, specifically annual report data sourced from the KKP for the specified period, excluding the year 2021, which was unavailable on the official website of the Ministry KKP as shown in Figure 1. Access to the dataset is provided via the following link: https://kkp.go.id/kategori/181-Laporan-Tahunan [32]. This research holds significant importance for several reasons. Firstly, by employing topic modeling techniques to analyze the annual reports of the KKP, the study aims to extract valuable insights and uncover latent patterns within the vast amount of textual data. Understanding the main topics and themes discussed in these reports is crucial for stakeholders, including policymakers, industry experts, and the public, as it provides a comprehensive overview of the Ministry's activities, priorities, and challenges over the specified period.

Figure 1. Cover of the 2022 KKP annual report: example from our dataset

The annual reports issued by the KKP serve as pivotal documents, encapsulating the outcomes and progress of KKP's endeavors throughout each fiscal year. These reports are not only instrumental in illustrating the implementation of KKP programs and activities but also serve as a cornerstone of accountability, providing stakeholders, including the President of the Republic of Indonesia and various other entities, with a transparent view of KKP's vision, mission, goals, and policy directives. The dataset utilized in this research spans from 2015 to 2022, with each annual report contributing unique insights into the evolution of KKP's strategies and achievements over time. However, analyzing these reports presents significant challenges due to their voluminous nature, with hundreds of pages and tens of thousands of words to sift through, necessitating efficient and effective methodologies to extract meaningful insights.

Table 1 offers a comprehensive breakdown of the dataset, detailing the number of pages and words contained within each annual report. Notably, the size of the reports varies year by year, with some reports spanning over 100 pages and containing tens of thousands of words. For instance, the 2019 report stands out as the largest in terms of word count, while the 2017 report is comparatively smaller. To address the formidable task of analyzing such extensive documents, this research leverages advanced computational tools and methodologies, specifically employing the LDA method. LDA provides a systematic and efficient means of uncovering underlying topics and themes within the annual reports, enabling a comprehensive understanding of KKP's strategic focus areas, challenges, and accomplishments across the years.

Table 1. Dataset KKP description

Document (year)	Pages	Words
Annual report (2015)	94	17.322
Annual report (2016)	77	14.782
Annual report (2017)	68	14.116
Annual report (2018)	120	34.171
Annual report (2019)	176	44.548
Annual report (2020)	171	42.975
Annual report (2022)	150	20.148

Given the complexity and breadth of the data, the adoption of LDA represents a significant advancement in the field of textual analysis. By automating the process of topic modeling, this research streamlines the analysis of annual reports, enabling researchers to extract insights more effectively and efficiently. Moreover, the findings generated through this approach offer valuable insights into the key priorities and initiatives undertaken by KKP, shedding light on the organization's contributions to marine affairs and fisheries management in Indonesia. Ultimately, this research not only enhances our understanding of KKP's activities but also underscores the importance of leveraging advanced computational techniques to navigate and derive insights from large volumes of textual data.

The systematic approach outlined in Figure 2, encompassing pre-processing, topic modeling, and post-processing stages, plays a pivotal role in harnessing the power of topic modeling to enhance the KKP publication process. The pre-processing stage serves as the foundation, laying the groundwork for subsequent analysis by ensuring that the textual data extracted from annual reports is cleaned and validated for suitability in the topic modeling process. By meticulously cleaning documents and validating the prepared data for compatibility with the LDA technique, this stage ensures the integrity and quality of the input data, thereby enhancing the accuracy of the topic modeling results. Moving into the topic modeling stage, the core of the process, the application of the LDA technique enables KKP to identify common topics and themes within the annual reports. By leveraging a probabilistic model, LDA uncovers latent structures within the textual data, providing insights into prevalent topics and concerns across different annual reports. This stage enables KKP to gain a deeper understanding of the content and focus areas of the reports, thereby informing the development of targeted communication strategies tailored to the specific needs and interests of diverse stakeholders. In the post-processing stage, the final phase of the topic modeling process, the focus shifts towards visualizing the results to facilitate interpretation and communication. By selecting relevant topics from the modeling results and visualizing them using techniques such as word clouds, KKP can effectively communicate key findings and insights derived from the topic modeling analysis. These visual representations serve as powerful tools for engaging stakeholders and conveying complex information in a clear and intuitive manner, thereby enhancing the relevance and impact of KKP's publications. The systematic approach to topic modeling outlined in Figure 2 enables KKP to leverage advanced computational techniques to enhance the effectiveness and relevance of its publications. By understanding prevalent topics and concerns within annual reports and tailoring its messaging accordingly, KKP can foster greater engagement and support for its initiatives, ultimately driving positive outcomes in the marine and fisheries sector.

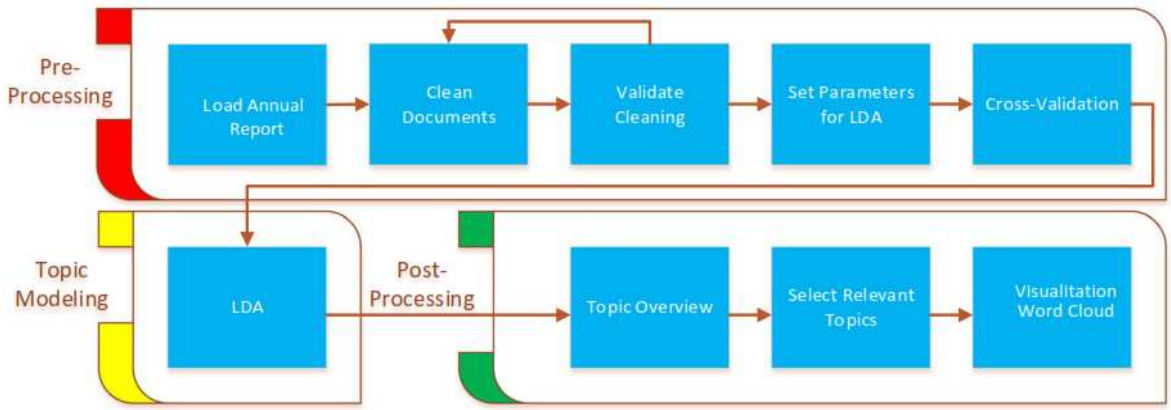

Figure 2. Flowchart of our KKP topic modeling system

4. RESULTS AND DISCUSSION

4.1. Fine-tuning LDA hyperparameters

Incorporating annual report data from 2015 to 2022 offers a comprehensive view of the KKP activities and initiatives over an extended period. By spanning multiple years, the analysis encompasses a wide range of policy directions, strategic objectives, and operational outcomes, providing a holistic understanding of KKP's evolution over time. This longitudinal approach enhances the robustness and reliability of the research findings, enabling researchers to identify overarching trends and patterns that may not be apparent within a single year's dataset.

The choice of the Python programming language, along with the Gensim package for LDA topic modeling, underscores the versatility and efficiency of modern computational tools in conducting sophisticated

text analysis tasks. Python's rich ecosystem of libraries and packages, coupled with its ease of use and flexibility, makes it a preferred choice for data scientists and researchers alike. The Gensim package, specifically designed for natural language processing tasks, offers powerful functionality for topic modeling, including LDA implementation with customizable parameters. The determination of the number of topics *(k)* in LDA modeling is a critical step in ensuring the quality and interpretability of the results. A well-chosen value for k balances the granularity of topic resolution with the coherence and interpretability of the generated topics. By tuning this parameter based on coherence scores, researchers can refine the topic modeling process and derive more meaningful insights from the textual data.

The coherence score serves as a vital metric for evaluating the quality of topics generated through LDA topic modeling. It assesses the degree of semantic coherence and interpretability of the topics, reflecting the extent to which the resulting topics are continuous, meaningful, and interrelated. Figure 3 illustrates that the coherence score tends to increase with the number of topics, indicating that discussions with a greater number of topics tend to exhibit higher coherence levels compared to those with fewer topics. To optimize the topic modeling process, parameter tuning is essential, involving the selection of an appropriate number of topics. In this research, the parameter tuning process involved iteratively assessing the coherence scores for different numbers of topics, ranging from 2 to 10. The model with 9 topics emerged as the most coherent, with a high coherence score of 0.4236. This indicates that the topics generated by this model exhibit a high degree of semantic coherence and relevance, making them more interpretable and meaningful. Table 2 (*see in Appendix*) provides insights into the distribution of words comprising each of the 11 best topics generated by the model. By examining the word distributions within each topic, researchers can gain a deeper understanding of the underlying themes and concepts captured by the model. This information is invaluable for interpreting and contextualizing the topics, enabling researchers to derive actionable insights and draw meaningful conclusions from the topic modeling analysis.

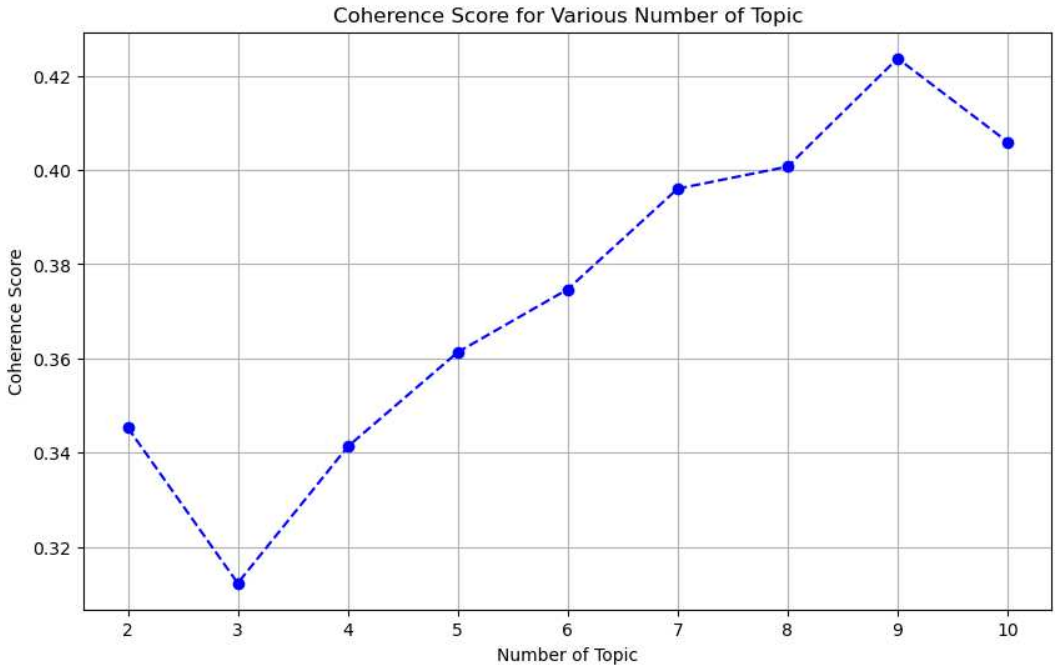

Figure 3. Coherence score of LDA tuning parameter

The analysis of annual reports from the KKP has yielded a diverse range of topics, each shedding light on critical aspects of fisheries and marine affairs in Indonesia. Topic 1 delves into the overarching theme of fisheries and marine sector development, highlighting the involvement of governmental ministries and efforts aimed at enhancing the quality and productivity of fisheries products. Similarly, topic 2 explores various facets of fish and marine-related activities, including area management, conservation efforts, and aquaculture practices, underscoring the multifaceted nature of marine resource utilization.

Meanwhile, topic 3 focuses on performance and development metrics at both provincial and national levels, emphasizing the importance of training initiatives, feed utilization, and infrastructure enhancement in

driving progress within the sector. The discussion surrounding topic 4 centers on fisheries resource monitoring, community engagement, and the economic implications of coastal resource utilization, highlighting the need for sustainable management practices to safeguard marine ecosystems and livelihoods. Topic 5 addresses support mechanisms for fishermen, including assistance programs, income generation initiatives, and infrastructure development, underscoring the pivotal role of supportive policies in bolstering the resilience of coastal communities.

Furthermore, topic 6 examines development activities across various locations, including investment strategies, infrastructure projects, and performance evaluation metrics, emphasizing the importance of strategic planning and resource allocation in fostering inclusive growth. Topic 7 delves into budgetary considerations and financial targets, particularly in eastern Indonesian districts, highlighting the significance of transparent budget allocation and accountability mechanisms in promoting equitable development outcomes. Similarly, topic 8 explores efforts to enhance the production value of fisheries commodities, optimize export opportunities, and achieve national economic objectives, underscoring the economic significance of the fisheries sector. Finally, topic 9 addresses quality improvement initiatives, community-based conservation efforts, and environmental preservation measures, particularly focusing on mangrove ecosystems in Java and other regions. Overall, the diverse array of topics derived from the analysis provides valuable insights into the challenges, opportunities, and policy imperatives shaping the fisheries and marine sector in Indonesia, facilitating informed decision-making and stakeholder engagement in sustainable resource management.

4.2. Extracting insights from KKP annual publications

This section delves into the exploration, comprehension, and evaluation of the results derived from the topic modeling process, leveraging the LDA technique. Through the utilization of the LDAVis and WordCloud packages, the analysis offers an enhanced, visually intuitive approach to gaining deeper insights into the identified topics. The visualization provided in Figure 4 offers a comprehensive overview of the distribution of the nine topics generated through LDA-based topic modeling. The LDAVis program illustrates the dispersion of these topics through bubbles, with each bubble representing a specific topic. Their relatively similar sizes suggest a balanced distribution of discussions across the annual reports of the KKP from 2015 to 2022.

Moreover, the accompanying bar chart in Figure 4 provides insights into the frequency of words present in the analysis. Notable recurring terms such as “fisheries,” “marine,” and “fish” underscore their significance within the corpus of annual reports, shedding light on the prevalent themes and focal points across the years. In Figure 5(a), the word cloud visualization further complements the analysis by highlighting the most frequently occurring words in the document set. The size of each word in the word cloud indicates its frequency of occurrence, with larger words signifying higher frequency. Notably, the prominence of certain terms in the word cloud aligns with their significance within the LDAVis visualization, reinforcing their importance in the annual reports.

Practically, these visualizations offer valuable insights for policymakers, researchers, and stakeholders involved in marine and fisheries management. By providing a visually intuitive representation of the topics and recurring terms within the annual reports, these visualizations facilitate the identification of key themes and discussions. This enables stakeholders to make informed decisions, develop targeted policies, and allocate resources effectively to address pertinent issues within the marine and fisheries sector. From a social perspective, the findings derived from the topic modeling analysis hold significant implications for coastal communities, fishermen, and other stakeholders reliant on marine resources for their livelihoods. By uncovering prevalent themes such as fisheries management, conservation efforts, and economic development, these findings can inform community-based initiatives, capacity-building programs, and sustainable resource management practices. This, in turn, contributes to the socio-economic well-being and resilience of coastal communities, fostering equitable and inclusive development.

Figure 5(b) shows the importance of these findings extends beyond the realm of academia and policy-making, resonating with broader societal concerns surrounding environmental sustainability, food security, and economic development. By shedding light on the challenges and opportunities within the marine and fisheries sector, these findings empower individuals, organizations, and communities to advocate for sustainable practices, promote environmental stewardship, and support the long-term health and vitality of marine ecosystems. By providing a nuanced understanding of the thematic distribution and recurring terms within the annual reports of the KKP, these findings enable informed decision-making, foster community engagement, and contribute to the advancement of sustainable development goals (SDGs) in Indonesia’s marine and fisheries sector.

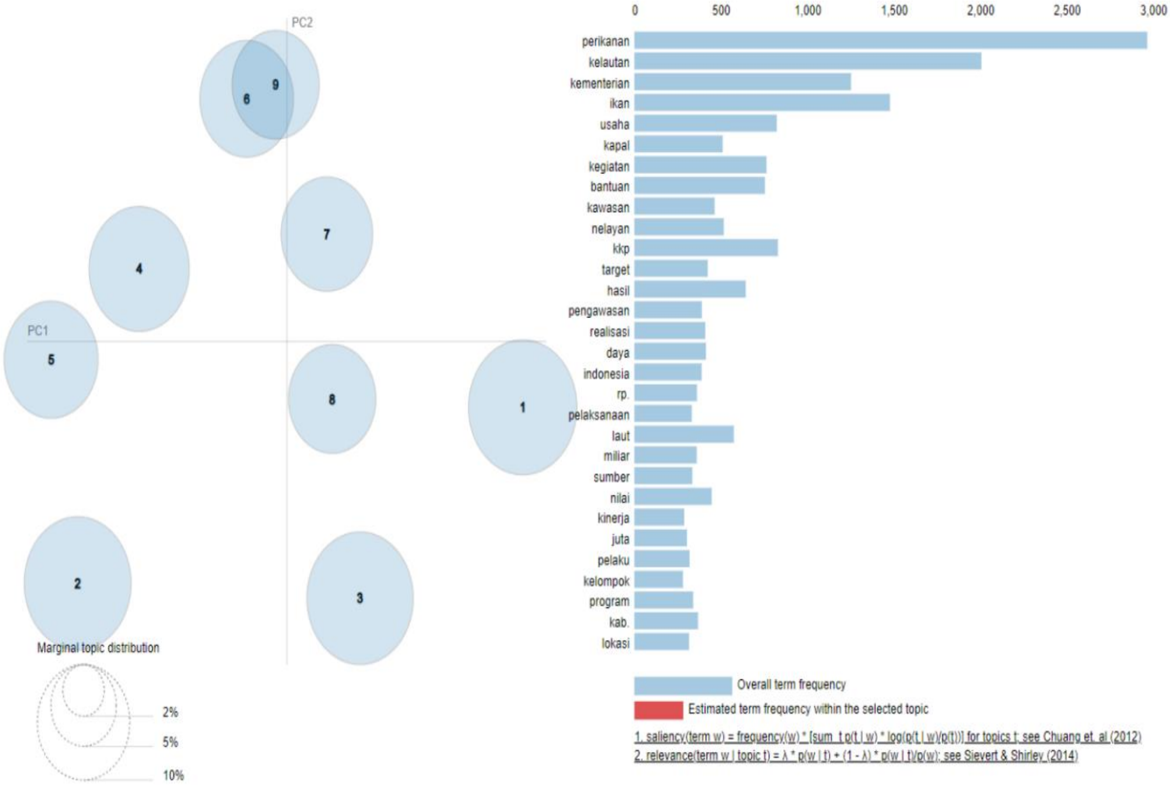

Figure 4. Visualization of 9 topics distribution across annual reports

(a)

(b)

Figure 5. Normalized word cloud visualization generated by lda-based topic modeling approach in;
(a) Indonesian and (b) English

4.3. Practical implication

The KKP in Indonesia plays a crucial role in managing and conserving the country's fisheries and marine resources. With the publication of its annual reports spanning from 2015 to 2022, KKP provides comprehensive insights into its activities, achievements, and challenges in advancing the fisheries and marine sector. Through the adoption of topic modeling techniques, these reports can be analyzed to extract valuable information and identify key themes relevant to KKP's mission and objectives. Practical implications can be drawn from each of the identified topics across the annual reports, offering actionable insights for KKP to enhance its policies, programs, and initiatives. Firstly, the emphasis on the fisheries and marine sector's development underscores the importance of cross-sectoral collaboration and innovation. KKP can leverage partnerships with ministries, industry stakeholders, and research institutions to drive forward initiatives aimed at improving fisheries products, enhancing value chain efficiency, and promoting sustainable practices.

Secondly, the focus on fish and marine-related aspects highlights the need for ecosystem-based management and conservation strategies. KKP can prioritize initiatives to strengthen marine protected areas, promote sustainable fishing practices, and mitigate the impact of climate change on marine ecosystems. By investing in habitat restoration, biodiversity conservation, and sustainable aquaculture, KKP can safeguard marine resources for future generations while supporting the livelihoods of coastal communities. Moreover, the emphasis on performance and development in fisheries and marine affairs underscores the importance of capacity building, infrastructure development, and policy support. KKP can prioritize training programs for fishery stakeholders, invest in modernizing fishing fleets and infrastructure, and enact policies to improve governance and regulatory frameworks. By enhancing the sector's performance and resilience, KKP can contribute to sustainable economic growth, food security, and poverty alleviation.

Furthermore, the acknowledgment of fisheries resource monitoring and community involvement highlights the importance of participatory approaches and stakeholder engagement in resource management. KKP can strengthen monitoring systems, promote community-based management initiatives, and empower local communities to participate in decision-making processes. By fostering collaboration and empowering stakeholders, KKP can enhance resource sustainability, promote social equity, and build resilience in coastal communities. Additionally, the recognition of budget targets and achievements in financial terms underscores the need for effective financial management and accountability. KKP can strengthen budget planning, execution, and monitoring processes to ensure transparent and efficient resource utilization. By prioritizing investments, implementing performance-based budgeting mechanisms, and fostering accountability, KKP can maximize the impact of budget allocations and achieve its development goals. The focus on enhancing the quality and performance of fisheries-related groups highlights the importance of capacity building, stakeholder collaboration, and environmental stewardship. KKP can invest in training programs, support community-led conservation initiatives, and enact policies to promote sustainable fisheries management. By empowering stakeholders, promoting environmental conservation, and fostering sustainable practices, KKP can enhance the resilience and productivity of fisheries-related activities. The practical implications derived from the KKP's annual reports provide valuable guidance for the Ministry to prioritize its actions, allocate resources effectively, and achieve its objectives. By addressing the key themes identified through topic modeling analysis, KKP can contribute to the sustainable development of Indonesia's fisheries and marine resources, promote social equity, and build resilience in coastal communities.

5. CONCLUSION

Ourstudy demonstrates the applicability of topic modeling for analyzing the annual reports of KKP Indonesia. Through rigorous evaluation and iteration, we achieved an optimal coherence value of 0.4236 with nine identified topics as the most coherent representation of the textual data. Each topic's distribution of ten words provides valuable insights into the interrelationships between various thematic elements present in the annual reports. Moreover, the significance of this study extends to critical sectors, including marine, coastal, and small island development, as well as fisheries, encompassing capture and aquaculture fisheries, along with the processing and marketing of fisheries products. Additionally, it holds relevance for the domains of supervision, research, and development within the fisheries sector. This comprehensive analysis not only illuminates the consistent focus on fisheries and marine-related initiatives but also underscores their importance for the sustainable development and management of marine resources, coastal areas, and small islands.The identified topics offer valuable insights for policymakers, researchers, and stakeholders, informing strategic decision-making, resource allocation, and policy formulation aimed at promoting sustainable development and stewardship of marine resources. Furthermore, the method, utilizing topic modeling techniques such as LDA, underscores the value of computational approaches for extracting meaningful insights from large textual datasets. The systematic evaluation of coherence scores and parameter tuning enhances the reliability and interpretability of the results, ensuring their relevance and applicability for diverse stakeholders.

Moving forward, the insights generated from this analysis can serve as a foundation for further research and policy development initiatives aimed at addressing key challenges facing the marine and fisheries sector in Indonesia. By leveraging advanced computational techniques and data-driven approaches, stakeholders can continue to refine their understanding of the sector, identify emerging trends, and implement targeted interventions to promote sustainable management practices and enhance the socio-economic well-being of coastal communities. In summary, the findings underscore the importance of topic modeling as a tool for analyzing and extracting insights from complex textual data, particularly in the context of the KKP annual reports. By uncovering latent patterns and relationships within the data, this research contributes to a more nuanced understanding of the dynamics shaping the marine and fisheries sector in Indonesia, ultimately supporting informed decision-making and sustainable development efforts in the region.

ACKNOWLEDGEMENTS

Author wish to acknowledge the research funding provided by Telkom University, which has been instrumental in supporting this study. The work of author was supported by the Ministry of Science and Technology, Taiwan, under Grant NSTC-111-2221-E-324-020, Grant NSTC-111-2622-E-324-002, Grant NSTC-112-2221-E-324-003-MY3, and Grant NSTC-112-2221-E-324-011-MY2. Rezzy Eko Caraka and Rung Ching Chen contributed equally as corresponding authors.

APPENDIX

Table 2. List of topics and top ten word distributions

Topic	10 Top word	Topic discussion
T1	*Perikanan – fisheries, Kelautan–Marine, Kementerian–Ministry, usaha-effort / business, hasil -result/yield, produk–product, pelaku-actor/participant, kapal-ship/vessel, sektor–sector, bidang - field/domain*	This topic would encompass discussions related to initiatives aimed at improving the fisheries and marine sector, with a focus on the involvement of government ministries and efforts to enhance the quality of fisheries products. It could involve strategies for sustainable fishing practices, regulatory measures for marine conservation, investment in infrastructure and technology, capacity building programs for fishermen, and initiatives to promote value-added products in the fisheries industry. The topic may also explore collaborations between government agencies, industry stakeholders, and local communities to address challenges and capitalize on opportunities in the fisheries and marine sector.
T2	*Ikan–fish, laut–sea, kawasan-area/region, pengelolaan–management, meningkatkan-improvement/enhancement, penangkapan-capture/fishing, konservasi–conservation, budidaya-cultivation farming, usaha-effort/enterprise, rangka-framework/skeleton*	This topic would encompass discussions related to various aspects of fish and marine management, including area management, conservation efforts, and the development of aquaculture activities. It would likely involve strategies for sustainable fisheries practices, marine resource conservation initiatives, and the promotion of responsible aquaculture practices. The topic may cover issues such as the establishment of marine protected areas, regulations for sustainable fishing practices, efforts to combat illegal fishing activities, and initiatives to enhance biodiversity and ecosystem resilience. Additionally, it could include discussions on the promotion of aquaculture as a means of supplementing wild fish stocks, supporting food security, and generating economic opportunities for coastal communities. Overall, this topic would highlight the importance of holistic approaches to managing and preserving fish and marine resources for present and future generations.
T3	*KKP, pembangunan–development, kinerja–performance, provinsi–province, nasional–national, orang-people/individuals, pakan–feed, pelatihan–training, bahan-material/ingredient, peaksanaan-treatment/handling*	This topic focuses on the performance and development of fisheries and marine affairs at both provincial and national levels, with a particular emphasis on key areas such as training, feed utilization, and infrastructure development. It would likely encompass discussions on initiatives aimed at improving the efficiency and effectiveness of fisheries management and marine resource utilization across different administrative levels. This could include strategies for capacity building and skill enhancement among stakeholders involved in the fisheries sector, as well as efforts to promote sustainable feed utilization practices to support aquaculture activities. Additionally, the topic may explore investments in infrastructure such as ports, harbors, and processing facilities to enhance the overall competitiveness and resilience of the fisheries and marine sectors. Overall, this topic sheds light on the progress and challenges in advancing fisheries and marine affairs, while highlighting the importance of collaborative efforts at both provincial and national levels.

Table 2. List of topics and top ten word distributions (*continued*)

Topic	10 Top word	Topic discussion
T4	*Pengawasan-supervision/monitoring, sumber–resource, masyarakat-community/society, daya-power/ability Indonesia–Indonesia, pemanfaatan-utilization/exploitation, wilayah-region/area, nomor–number, pesisir–coastal, ekonomi-economy*	This topic delves into the multifaceted aspects of fisheries resource management, community engagement, utilization of coastal areas, and their broader implications on the Indonesian economy. It encompasses discussions on the systematic monitoring and assessment of fisheries resources to ensure sustainability and mitigate overexploitation. Additionally, it explores the involvement of local communities in fisheries management processes, emphasizing the importance of participatory approaches and indigenous knowledge systems. Moreover, the topic examines the utilization of coastal areas for various purposes, including fishing activities, aquaculture, tourism, and conservation efforts, and evaluates their socio-economic impacts on local communities and the national economy. Through an integrated approach, this topic aims to provide insights into the complex interactions between fisheries management, community dynamics, coastal development, and economic outcomes, thereby informing policy decisions and management strategies for sustainable marine resource utilization.
T5	*Bantuan-assistance/aid, realisasi-realization/implementation, sistem–system, nelayan-fisherman/fisherfolk, sarana-facility/infrastructure, capaian-achievement/attainment, pendapatan-income/revenue, prasarana-infrastructure/facilities, penerapan-application/implementation, kategori-category/classification*	This topic addresses the comprehensive support systems designed to aid fishermen and bolster fisheries activities, encompassing discussions on various forms of assistance, achievements in support programs, and the socio-economic impacts on fishing communities. It examines the range of assistance provided to fishermen, including financial support, training programs, technological advancements, and access to markets, aimed at enhancing their livelihoods and promoting sustainable fishing practices. Additionally, it evaluates the effectiveness of support systems in achieving their intended goals, such as increased productivity, improved living standards, and enhanced resilience to environmental challenges. Furthermore, the topic explores the socio-economic impacts of these support systems on fishing communities, including changes in income levels, livelihood diversification, and improvements in infrastructure facilities such as harbors, landing sites, and processing centers. By analyzing the assistance provided, achievements made, and socio-economic outcomes, this topic provides valuable insights into the efficacy of support systems in fostering sustainable fisheries and promoting the well-being of fishing communities.
T6	*Kegiatan–activity, lokasi–location, kota–city, rencana–plan, investasi–investment, mesin-machine/engine, pulau-pulau–islands, pembangunan–development, pengembangan-development/expansion, indikator-indicator*	This topic explores the activities and development plans across various locations, with a particular focus on investment and infrastructure development initiatives on islands. It encompasses discussions on the strategies and initiatives aimed at promoting sustainable development and enhancing economic growth in diverse geographical settings. Specifically, it examines the investment plans and infrastructure development projects targeting islands, which play a crucial role in supporting local economies, fostering connectivity, and improving living standards. Moreover, the topic emphasizes the importance of performance indicators in evaluating the effectiveness and progress of development plans, providing insights into key metrics used to measure success and identify areas for improvement. By analyzing investment strategies, infrastructure development efforts, and performance monitoring mechanisms, this topic contributes to a better understanding of regional development dynamics and informs decision-making processes aimed at achieving sustainable and inclusive growth across various locations.
T7	*Target–target, RP-rupiah (Indonesian currency), miliar–billion, kab.-kabupaten (district), kabupaten–district, mencapai-reach/achieve, timur–east, juta-million anggaran–budget, penyusunan-arrangement/preparation*	This topic delves into the budget targets and achievements in financial terms, particularly within the context of development initiatives across various districts in eastern Indonesia. It encompasses discussions on the allocation of budgets, financial planning strategies, and the utilization of funds to support development projects and initiatives in the region. Furthermore, the topic examines the achievements and outcomes of these financial allocations, assessing their impact on local communities, infrastructure development, and socio-economic progress. With a specific focus on districts in eastern Indonesia, the topic sheds light on the unique challenges, opportunities, and priorities faced by these regions in their development efforts. By analyzing budget targets, financial achievements, and their implications for regional development, this topic provides valuable insights into the effectiveness of financial management practices and informs future decision-making processes aimed at promoting sustainable and inclusive development in eastern Indonesia.

Table 2. List of topics and top ten word distributions (*continued*)

Topic	10 Top word	Topic discussion
T8	*Nilai–value, ikan–fish, garam–salt, paket–package, produksi–production, tujuan-destination/goal, ekspor–export, peningkatan-increase/improvement, negara-country/nation, proses-process*	This topic delves into discussions surrounding the production value of fish and salt, alongside efforts aimed at increasing exports to bolster the achievement of the country's economic objectives. It encompasses analyses of the production value of fish and salt, highlighting their significance as key commodities in the national economy. Furthermore, the topic explores strategies and initiatives designed to boost exports of fish and salt, considering factors such as market demand, trade regulations, and supply chain logistics. Additionally, it examines the role of exports in contributing to the country's economic goals, including foreign exchange earnings, job creation, and GDP growth. By evaluating efforts to enhance export-oriented fisheries and salt production, this topic provides insights into the potential benefits and challenges associated with promoting economic growth while ensuring sustainability and competitiveness in the global market.
T9	*Kelompok–group, mutu–quality, kinerja–performance, barat–west, dilaksanakan-implemented/carried out, selatan–south, penanganan-handling/management, kampung–village, mangrove–mangrove, jawa-Java*	This topic explores strategies aimed at improving the quality and performance of diverse stakeholders within the fisheries sector, with a particular emphasis on initiatives related to handling and environmental preservation, including mangrove conservation efforts in various regions, particularly in Java. It encompasses discussions on interventions aimed at enhancing the skills, knowledge, and capabilities of individuals and groups involved in fisheries activities, such as fishermen, aquaculture practitioners, and coastal communities. Furthermore, the topic examines the importance of environmental preservation, with a specific focus on mangrove ecosystems, which play a crucial role in supporting marine biodiversity, coastal protection, and sustainable fisheries. It analyzes strategies for mangrove conservation, restoration, and sustainable management, considering the socio-economic benefits and ecological significance of these efforts. By emphasizing the nexus between quality improvement, performance enhancement, and environmental preservation, this topic provides insights into holistic approaches to fostering sustainable development and resilience within the fisheries sector, particularly in the context of mangrove conservation in Java and beyond.

REFERENCES

[1] F. A. Hudaefi, "How does Islamic fintech promote the SDGs? Qualitative evidence from Indonesia," *Qualitative Research in Financial Markets*, vol. 12 no. 4, pp. 353-366, 2020, doi: 10.1108/QRFM-05-2019-0058.

[2] F. A. Hudaefi and N. Heryani, "The practice of local economic development and *Maqāṣid al-Sharī'ah*: Evidence from a *Pesantren* in West Java, Indonesia," *International Journal of Islamic and Middle Eastern Finance and Management*, vol. 12 no. 5, pp. 625-642, 2019, doi: 10.1108/IMEFM-08-2018-0279.

[3] F. A. Hudaefi and K. Noordin, "Harmonizing and constructing an integrated *Maqāṣid al-Sharī'ah* index for measuring the performance of Islamic banks," *ISRA International Journal of Islamic Finance*, vol. 11, no. 2, pp. 282-302, 2019, doi: 10.1108/IJIF-01-2018-0003.

[4] C. J. De La Torre, D. Sánchez, I. Blanco, and M. J. Martin-Bautista, "Text mining: techniques, applications, and challenges," *International journal of uncertainty, fuzziness and knowledge-based systems*, vol. 26, no. 04, pp. 553–582, 2018, doi: 10.1142/S0218488518500265.

[5] H. Hashimi, A. Hafez, and H. Mathkour, "Selection criteria for text mining approaches," *Comput Human Behav*, vol. 51, pp. 729–733, Oct. 2015, doi: 10.1016/j.chb.2014.10.062.

[6] R. E. Caraka *et al*., "Connectivity, sport events, and tourism development of Mandalika's special economic zone: A perspective from big data cognitive analytics," *Cogent Business and Management*, vol. 10, no. 1, 2023, doi: 10.1080/23311975.2023.2183565.

[7] A. Shabilla and W. S. Nugroho, "Business and Economics Conference in Utilization of Modern Technology Magelang," *Business and Economics Conference in Utilization of Modern Technology*, 2020.

[8] R. E. Caraka *et al*., "Strategic insights for MSMEs: navigating the new normal with big data and business analytics," *Journal of Asia Business Studies*, 2023, doi: 10.1108/JABS-10-2022-0354.

[9] F. A. Hudaefi, R. E. Caraka, and H. Wahid, "Zakat administration in times of COVID-19 pandemic in Indonesia: a knowledge discovery via text mining," *International Journal of Islamic and Middle Eastern Finance and Management*, vol. 15, no. 2, pp. 271-286, 2022, doi: 10.1108/IMEFM-05-2020-0250.

[10] A. Usai, M. Pironti, M. Mital, and C. A. Mejri, "Knowledge discovery out of text data: a systematic review via text mining," *Journal of Knowledge Management*, vol. 22, no. 7, pp. 1471-1488, 2018, doi: 10.1108/JKM-11-2017-0517.

[11] S. Lee, J. Song, and Y. Kim, "An empirical comparison of four text mining methods," Journal of Computer Information Systems, vol. 51, no. 1, pp. 1–10, 2010.

[12] M. Jiang *et al*., "A variety of text mining technology and tools research," in *2014 International Conference on Mechatronics, Electronic, Industrial and Control Engineering, MEIC 2014*, 2014, pp. 918–921, doi: 10.2991/meic-14.2014.203.

[13] K. Lo, F. Ramos, and R. Rogo, "Earnings management and annual report readability*," Journal of Accounting and Economics,* vol. 63, no. 1, pp. 1–25, Feb. 2017, doi: 10.1016/j.jacceco.2016.09.002.

[14] K. L. Sumathy and M. Chidambaram, "Text Mining: Concepts, Applications, Tools and Issues-An Overview," *International Journal of Computer Applications (0975 – 8887)*, vol. 80, no. 4, pp. 29-32, Oct. 2013.

[15] I. A. Aditya *et al*., "Understanding service quality concerns from public discourse in Indonesia state electric company," *Heliyon*, vol. 9, no. 8, Aug. 2023, doi: 10.1016/j.heliyon.2023.e18768.

[16] E. Nylander, A. Fejes, and M. Milana, "Exploring the themes of the territory: a topic modelling approach to 40 years of publications in International Journal of Lifelong Education (1982–2021)," *International Journal of Lifelong Education*, vol. 41, no. 1, pp. 27–44, 2022, doi: 10.1080/02601370.2021.2015636.
[17] L. Muchene and W. Safari, "Two-stage topic modelling of scientific publications: A case study of University of Nairobi, Kenya," *PLoS One*, vol. 16, no. 1, Jan. 2021, doi: 10.1371/journal.pone.0243208.
[18] S. I. Nikolenko, S. Koltcov, and O. Koltsova, "Topic modelling for qualitative studies," *Journal of Information Science*, vol. 43, no. 1, pp. 88–102, Feb. 2017, doi: 10.1177/0165551515617393.
[19] K. R. Canini, L. Shi, and T. L. Griffiths, "Online inference of topics with latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 5, pp. 65–72, Apr. 2009.
[20] S. Syed and M. Spruit, "Exploring Symmetrical and Asymmetrical Dirichlet Priors for Latent Dirichlet Allocation," *International Journal of Semantic Computing*, vol. 12, no. 3, pp. 399–423, 2018, doi: 10.1142/S1793351X18400184.
[21] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The Author-Topic Model for Authors and Documents," *arXiv preprint arXiv:1207.4169*, 2012.
[22] H. W. Cho, "Editorial: Topic modeling," *Osong Public Health and Research Perspectives*, vol. 10, no. 3, pp. 115–116, 2019, doi: 10.24171/j.phrp.2019.10.3.01.
[23] D. M. Blei and J. D. Lafferty, "Dynamic Topic Models," *Proceedings of the 23rd international conference on Machine learning*, Jun. 2006, pp. 113–120, doi: 10.1145/1143844.1143859.
[24] P. Kherwa and P. Bansal, "Topic Modeling: A Comprehensive Review," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 7, no. 24, pp. 1–16, 2020, doi: 10.4108/eai.13-7-2018.159623.
[25] R. Alghamdi and K. Alfalqi, "A Survey of Topic Modeling in Text Mining," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 6, no. 1, pp. 147-153, 2015.
[26] Y. Lu, Q. Mei, and C. X. Zhai, "Investigating task performance of probabilistic topic models: An empirical study of PLSA and LDA," *Inf Retr Boston*, vol. 14, no. 2, pp. 178–203, Apr. 2011, doi: 10.1007/s10791-010-9141-9.
[27] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of machine Learning research*, pp. 993-1022, Jan. 2003.
[28] J. C. Campbell, A. Hindle, and E. Stroulia, "Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data," *Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data*, pp. 139-159, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.
[29] [K. Watanabe and A. Baturo, "Seeded sequential lda: A semi-supervised algorithm for topic-specific analysis of sentences," *Social Science Computer Review*, vol. 42, no. 1, pp. 224–248, May 2023. doi:10.1177/08944393231178605
[30] M. Seyfert and I. Viola, "Dynamic word clouds," *Proceedings of the 33rd Spring Conference on Computer Graphics*, May 2017, doi:10.1145/3154353.3154358.
[31] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, "Word Cloud Explorer: Text Analytics Based on Word Clouds," *2014 47th Hawaii International Conference on System Sciences*, Waikoloa, HI, USA, 2014, pp. 1833-1842, doi: 10.1109/HICSS.2014.231.
[32] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," *EMNLP 2011-Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, no. 2, pp. 262–272, 2011.

BIOGRAPHIES OF AUTHORS

Ahmad Marzuqi received the Bachelor of Informatics degree from the School of Computing, Telkom University, in 2021. He is currently taking a Master's degree in Management at the University of Indonesia. His academic pursuits are fueled by interest in big data, data analysis, and artificial intelligence, reflecting his interest in exploring and advancing knowledge in these fields. He can be contacted at email: ahmad.marzuqi21@ui.ac.id.

Rezzy Eko Caraka currently serves as a researcher at the Research Center for Data and Information Sciences within the Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Indonesia, a position he has held since February 2022. Prior to this role, he served as a Post-doctoral Researcher with the Department of Statistics at Seoul National University from 2019 to December 2021, and later as a Post-Doctoral Researcher with the Department of Nuclear Medicine at Seoul National University Hospital from January 2021 to January 2022. He also assumed the role of Research Assistant Professor at the Department of Statistics, Seoul National University, from January to April 2022. In addition to his research roles, he holds several academic positions. He has been an adjunct lecturer with the Faculty of Economics and Business at Universitas Indonesia since 2021, an adjunct lecturer at the Graduate School, Department of Statistics at Padjadjaran University since 2021, and a senior research fellow with the Department of Mathematics at Ulsan National Institute of Science and Technology, South Korea, since 2022. Notably, he assumed the position of senior lecturer at Telkom University starting in 2024. He has been recognized for his contributions to the field, being acknowledged as a top 2% researcher in AI and machine learning by Stanford University. His diverse research interests encompass statistics, large-scale optimization, machine learning, big data analytics, data science, and sustainable development goals. He can be contacted at email: rezzy.eko.caraka@brin.go.id or rezzy.eko@ui.ac.id or rezzyekocaraka@telkomuniversity.ac.id.

Prana Ugiana Gio is the founder of STATCAL (statistical software) https://statcal.com/ and content creator on the Youtube channel: STATKOMAT (programming statistics). He is a lecturer at the Department of Mathematics, Universitas Sumatera Utara. His field of study is building web-based applications using R and Javascript, probability distribution modeling, Monte Carlo simulation, and Bayesian. He has published dozens of books related to programming and statistics. He can be contacted at email: prana@usu.ac.id.

Rung Ching Chen received the B.S. degree from the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, in 1987, the M.S. degree from the Institute of Computer Engineering, National Taiwan University of Science and Technology, in 1990, and the Ph.D. degree in computer science from the Department of Applied Mathematics, National Chung Hsing University, in 1998. He is currently a Distinguished Professor with the Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan. He is listed in the top 2% scientists worldwide in A.I. by Stanford University. His research interests include network technology, pattern recognition, knowledge engineering, the internet of things, data analysis, and artificial intelligence. He can be contacted at email: crching@cyut.edu.tw.

Maengseok Noh was born in Busan, South Korea, in 1973. He received the B.S., M.S., and Ph.D. degrees from the Department of Statistics, Seoul National University, in 1996, 1998, and 2005, respectively. His thesis was on analyzing binary data and robust modeling via hierarchical likelihood. Since 2006, he has been a Professor with the Department of Statistics, Pukyong National University, Busan. His current research interests include the application and software development for hierarchical generalized linear models, methodology development for zero-inflated poisson model with spatial correlation, and hierarchical approach non-Gaussian factor analysis. He can be contacted at email: msnoh@pknu.ac.kr.

Bens Pardamean has over thirty years of global experience in information technology, bioinformatics, and education. His professional experience includes being a practitioner, researcher, consultant, entrepreneur, and lecturer. He currently holds a dual appointment as Director of Bioinformatics and Data Science Research Center (BDSRC) | AI Research and Development Center (AIRDC) and Professor of Computer Science at Bina Nusantara (BINUS) University in Jakarta, Indonesia. He earned a Doctoral degree in Informatics Research from University of Southern California (USC), as well as a Master's degree in Computer Education and a Bachelor's degree in Computer Science from California State University at Los Angeles (USA). He can be contacted at email: bpardamean@binus.edu.