TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 5, October 2026, pp. 1550~1560
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i5.27674

Exploration of thesis topic trends of students majoring in informatics and computer engineering with LDA method

Ruslan, Rezki Nurul Jariah S. Intam, Sasmita, Dewi Fatmarani Surianto, Andi Akram Nur Risal, Nur Azizah Eka Budiarti
Department of Computer Engineering, Faculty of Engineering, University of Makassar, South Sulawesi, Indonesia

Article Info

Article history:

Received Nov 18, 2025
Revised Jun 5, 2026
Accepted Jun 30, 2026

Keywords:

Informatic and computer engineering
Latent Dirichlet allocation
Thesis
Topic exploration
Topic modeling
Web-based systems

ABSTRACT

The selection of a thesis topic is very important because it determines the focus of the research and its contribution to knowledge. However, many students find it difficult to choose a topic that suits their interests and expertise. This study models the thesis topics of students in the Department of Informatics and Computer Engineering (JTIK) using the latent Dirichlet allocation (LDA) method, with term frequency–inverse document frequency (TF–IDF) as the model input. The data set includes 969 thesis titles from 2009 to 2024. The optimized LDA model identifies 17 main topics by adjusting parameters such as the number of topics, alpha, and beta. The best coherence value (0.7431) is achieved with alpha = 0.81, beta = 0.01, and 17 topics. The dominant themes included information system development, computer networks, and multimedia, reflecting the main research areas of JTIK. In addition, a web-based system was developed and integrated with the best model to help students identify relevant topics and find thesis references. This study demonstrates the effectiveness of topic modeling in higher education and provides insights into academic research trends.

Corresponding Author:

Dewi Fatmarani Surianto
Department of Computer Engineering, Faculty of Engineering, University of Makassar
South Sulawesi, Indonesia
Email: dewifatmaranis@unm.ac.id

1. INTRODUCTION

A thesis is a written work completed by students as a requirement for obtaining a bachelor's degree and reflects their specialization in a particular field of study [1]-[4]. Every year, the Department of Informatics and Computer Engineering (JTIK), Faculty of Engineering, Makassar State University, graduates' students from two study programs, namely Computer Engineering (TEKOM) and Informatics and Computer Engineering Education (PTIK), which have different research characteristics and specialization focuses. These differences contribute to the diversity of thesis themes and research approaches produced by students.

The selection of a thesis topic is an important academic stage because it determines the direction and contribution of the research [5], [6]. However, students often experience difficulties due to limited references, mismatched expertise, and insufficient guidance [7]. Therefore, topic modeling can help students identify research trends and align thesis topics with their interests, competencies, and academic goals [8].

Topic modeling is a probabilistic approach used to identify hidden topics in a collection of documents [9]. One of the most widely used methods is latent Dirichlet allocation (LDA), which extracts latent topics through groups of related keywords [10]. LDA has been widely applied in various domains, including agricultural documents, social media data, research trend analysis, and student thesis titles [11]-[15]. However, previous studies generally produced relatively low coherence values, particularly on short-text

datasets such as thesis titles. Alternative approaches such as latent semantic analysis (LSA) and BERTopic have also been explored, but topic consistency and interpretability remain challenging [16], [17]. Therefore, this study proposes the optimization of the LDA method for modeling JTIK student thesis topics to improve topic coherence and representation quality. Improved coherence is expected to generate more accurate and consistent topic interpretation, provide insights into research trends in the JTIK academic environment, and assist students in determining relevant thesis topics.

2. METHOD

The research methodology involves steps or a series of processes to model topics in JTIK student theses at Makassar State University. These steps include data collection, data preprocessing, implementation of topic modeling techniques, and evaluation of results. Further details can be seen in Figure 1.

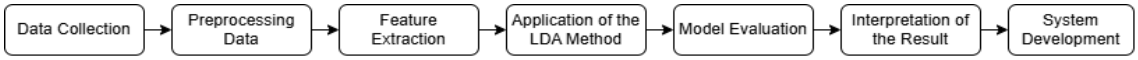


Figure 1. Research stages

2.1. Data collection

The initial stage of this research involved collecting thesis titles of students from the JTIK, Makassar State University, from 2009 to 2024. The data were obtained from the JTIK e-scription website, which stores published student theses. Examples of the collected data are presented in Table 1.

Based on a statistical analysis of the collected data, the average document length is approximately 15.28 tokens per title, with a minimum of 3 tokens and a maximum of 32 tokens. A statistical distribution of the number of tokens per title is shown in Figure 2. As shown in Figure 2, most thesis titles range between 10 and 20 tokens, with the peak centered around the average value. The vertical line indicates that the mean document length is approximately 15.28 tokens per title.

Table 1. Example of data used

No.	Title
1	Study of generator protection systems at PT Semen Tonasa Power Plant
2	Development of a product sales information system at PD Putra Syarif in Soppeng Regency

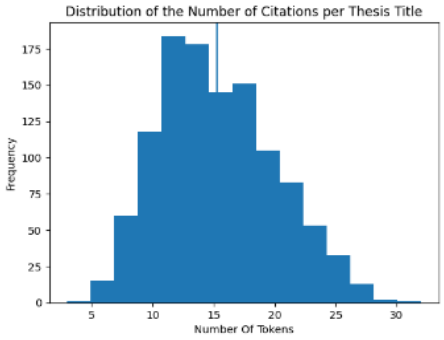


Figure 2. Document length statistics

2.2. Data preprocessing

The next step is data preprocessing to remove redundant information and improve the model learning process [17]. The preprocessing stages include case folding, stopword removal, and tokenization to prepare JTIK thesis titles for further analysis. Examples of the preprocessing results are presented in Table 2.

Table 2. Example of preprocessing results

No.	Title
1	[‘system’, ‘protection’, ‘factory’, ‘steam’, ‘PLTU’, ‘PT’]
2	[‘development’, ‘system’, ‘information’, ‘product’, ‘pd’, ‘son’, ‘syarif’]

2.3. Feature extraction

At the feature extraction stage, this study uses term frequency–inverse document frequency (TF-IDF) to represent documents because it considers both word frequency and word importance within the corpus [18]. Previous studies have shown that TF-IDF is effective in highlighting meaningful words and improving document representation quality for topic modeling tasks [19]. Table 3 presents an example of TF, IDF, and TF-IDF values used for word weighting in this study. Each word has a TF, IDF, and TF-IDF value that reflects its importance in the document. While all words have the same TF value, the IDF and TF-IDF values vary depending on word frequency in the corpus. Words that appear less frequently, such as "protection," "generator," and "steam," obtain higher TF-IDF values, indicating stronger relevance, whereas common words such as "system" produce lower TF-IDF values.

Table 3. Example of feature extraction results

ID	Term	TF	ID	TF-IDF
10	System	0.16	1.18	0.20
16	Protection	0.16	6.88	1.15
17	Generator	0.16	6.88	1.15
18	Steam	0.16	6.88	1.15

2.4. Application of the LDA method

At this stage, the LDA method is applied to model topics in JTIK student theses. LDA will identify latent or hidden topics in thesis documents. The process of finding hidden topics is done by grouping words that have semantic relationships based on probability distribution calculations. The formula used in the implementation of the LDA technique is in (1) [10]:

$$P(\beta_{1:k}, 0_{1:D}, Z_{1:D,1:N}, W_{1:D,1:N}) = \prod^{K} p(\beta_k|n) \prod_{d=1}^{D} p(0_d|a) \left(\prod_{n=1}^{N} p\left(z_{d,n \mid 0_d}\right) p\left(W_{d,n} \middle| \beta_{1:k,}\, Z_{d,n}\right)\right. \quad (1)$$

$\beta k\theta DZDNWDN,$ Where the (1) describes the combined probability of topic parameters $(1{:}\,\beta 1{:}\,k)$, document topic distribution $(1{:}\,\theta 1{:}\,D)$, topic word distribution in documents $(1{:},1{:}\,Z1{:}\,D, 1{:}\,N)$, and actual words in documents $(1{:},1{:}\,W1{:}\,D, 1{:}\,N)$ in the document topic probabilistic model.

2.5. Hyperparameter adjustment

At this stage, hyperparameter adjustment is performed to optimize the performance of the LDA model. The adjustment focuses on determining the optimal values of the alpha, beta, and number of topics parameters based on the coherence score. The combination of hyperparameters that produces the highest coherence score is selected as the optimal configuration for the LDA model.

2.6. Model evaluation

This stage evaluates the model using coherence values to measure the semantic relationship between words within each topic. Higher coherence values indicate more consistent and meaningful topics. The topic modeling results are then visualized using word distribution graphs.

2.7. Interpretation of results

At this stage, the resulting topics are analyzed and interpreted based on the keywords associated with each topic. Topic visualization is conducted using the PyLDAvis library to examine the distribution of terms and the interrelationships among the identified topics. The resulting topic structures are further evaluated to identify the dominant themes and thematic patterns within the thesis titles.

2.8. System development

At this stage, a web-based system is developed to present the topic modeling results. Flask is used to implement the interactive interface for displaying the identified topics and thesis titles. The development follows Agile principles through an iterative process to support continuous system improvement.

3. RESULTS AND DISCUSSION

3.1. Test results

Topic modeling in this study uses the LDA method as the main method. As a form of evaluation and comparison, this study also applies other topic modeling methods, namely LSA, non-negative matrix factorization (NMF), and BERTopic. The results of comparing these methods can be seen in Table 4.

Table 4. Method comparison

Method	Number of Topics	Coherence
LDA	17	0.7431
LSA	17	0.3860
NMF	17	0.6591
BERTopic	17	0.6512

The comparison results show that LDA achieved the highest coherence value of 0.7431 with 17 topics, outperforming NMF (0.6591), BERTopic (0.6512), and LSA (0.3860). These findings indicate that LDA is more effective for modeling short-text thesis title data and was therefore selected as the main method in this study. Furthermore, a comparison between the original Bag-of-Words-based LDA [20] and LDA with TF-IDF weighting was conducted to evaluate the effect of document representation on topic coherence. The comparison results are presented in Table 5.

The comparison results show that LDA with TF-IDF weighting achieved a higher coherence value (0.7431) than the original Bag-of-Words-based LDA (0.4779) using the same number of topics 17. This indicates that TF-IDF improves the quality and coherence of topics generated from thesis title data. Based on these findings, topic modeling was performed using LDA with parameter optimization of the number of topics 2–19, alpha, and beta, resulting in 450 model combinations. The parameter ranges were determined according to the characteristics of short-text thesis titles and previous studies suggesting that smaller alpha and beta values produce more focused and coherent topics in sparse documents [21], while a small to medium number of topics helps avoid topic fragmentation [22]. The parameter testing results are presented in Table 6.

Table 6 shows the test results of 450 LDA models, where the best model was obtained with parameters $\alpha = 0.81$, $\beta = 0.01$, and 17 topics using TF-IDF input. This model achieved the highest coherence value of 0.7431, showing a significant improvement compared to other parameters [23]. Based on this best model, further tests were conducted to analyze the effect of β while keeping the number of topics and α constant, with the results shown in Table 7.

Based on Table 7, the smallest beta value (0.01) produced more specific topics with focused words, resulting in the highest coherence of 0.7431. As the beta value increased, the topics became more scattered and less specific, decreasing coherence; at a beta value of 0.81, coherence dropped to 0.3628, indicating overly general topics. Next, the effect of alpha value variation was tested while keeping the number of topics and beta constant, as shown in Table 8.

Table 5. Comparison results of coherence between original LDA and LDA + TF-IDF

Scenario	Number of Topics	Coherence
Original LDA	17	0.4779
LDA +TF-IDF	17	0.7431

Table 6. Parameter adjustment results

No.	Number of topics	Alpha	Beta	Coherence
1	2	0.01	0.01	0.3975
2	2	0.01	0.21	0.3817
...	...	...	...	...
321	17	0.81	0.01	0.7431

Table 7. Beta value test results

No.	Number of topics	Alpha	Beta	Coherence
1	17	0.81	0.01	0.7431
2	17	0.81	0.21	0.6843
…	…	…	….	…
5	17	0.81	0.81	0.3628

Table 8. Alpha value testing results

No.	Number of topics	Alpha	Beta	Coherence
1	17	0.01	0.01	0.5882
2	17	0.21	0.01	0.6380
…	…	…	…	…
5	17	0.81	0.01	0.7431

Based on Table 8, topic coherence increases as the alpha value rises, reaching the highest coherence of 0.7431 at $\alpha = 0.81$. This indicates that higher alpha values produce more evenly distributed and coherent topics. Furthermore, the effect of varying the number of topics while keeping alpha and beta constant is presented in Table 9.

Table 9. Test results for number of topics

No.	Number of Topics	Alpha	Beta	Coherence
1	2	0.81	0.01	0.4246
2	3	0.81	0.01	0.5442
...	...	...	...	...
16	17	0.81	0.01	0.7431

Based on Table 9, coherence values increased with the number of topics and reached the highest value of 0.7431 at 17 topics, which was determined as the optimal number before coherence began to decrease [24]. These results indicate that the combination of a high alpha value, low beta value (0.01), and an optimal number of topics produces more specific, relevant, and semantically consistent topics for sparse thesis title data.

3.2. Interpretation and analysis of topic modeling

After obtaining the model with the highest coherence, topic labeling was performed based on dominant keywords and thesis titles with the highest probability in each topic and supported by the inter-topic distance map visualization, which in Figure 3 displays 17 topics and 30 relevant terms, where topic 1 has proximity to topics 6, 14, and 15 (±6.3%) with the model's main keywords, improving, learning, and use, thus labeled as "technology-based learning model" [25].

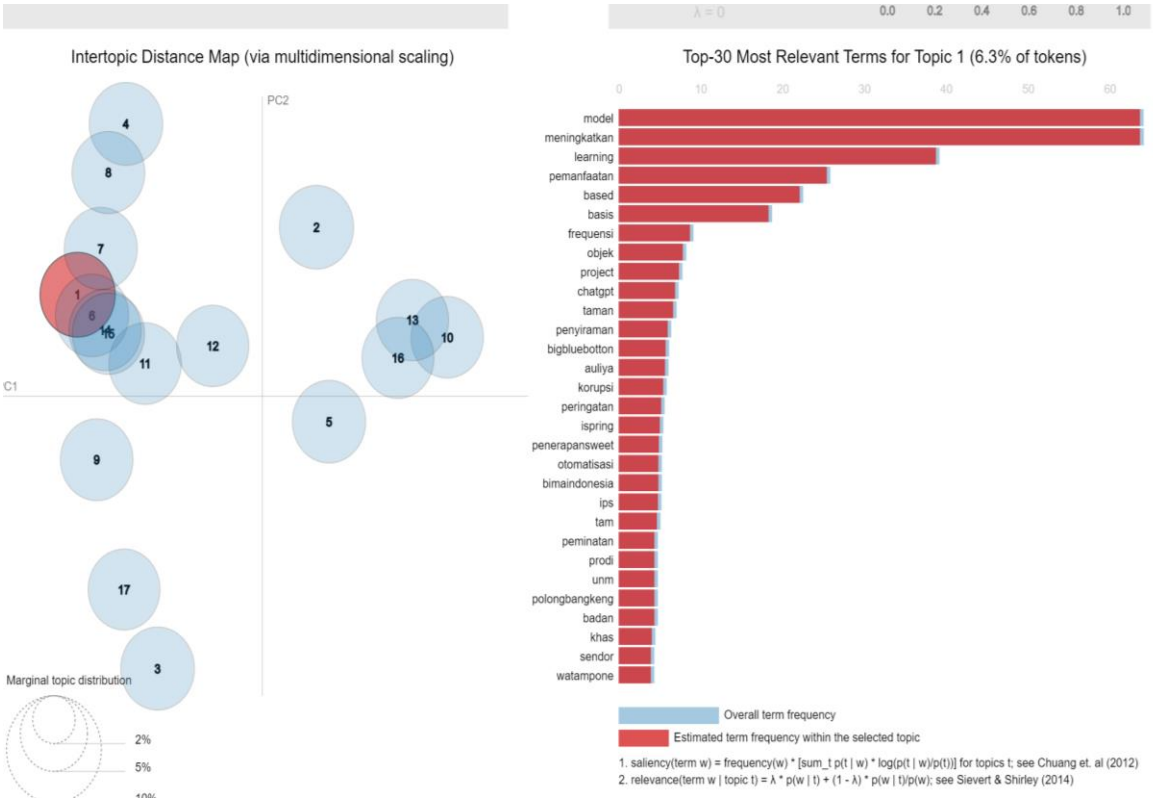


Figure 3. Topic distance map visualization

Keywords from topics 1–17 were used to assign topic labels (Figure 4). The labels were determined through keyword analysis and visualization interpretation to represent the main research themes identified by the LDA model. The resulting 17 topics were grouped into categories related to technology, information systems, and learning, showing distinct research focuses despite some keyword overlap caused by the short-text nature of thesis titles. To validate the semantic appropriateness of the labels, two experts evaluated each topic using the percentage agreement method with a binary scale. The evaluation results are presented in Table 10.

Figure 4. Keywords for each topic

Table 10. Results of expert agreement evaluation on topic labels

Number of items	Items agreed	Items disagreed	Percentage agreement (%)
17	13	4	76.47

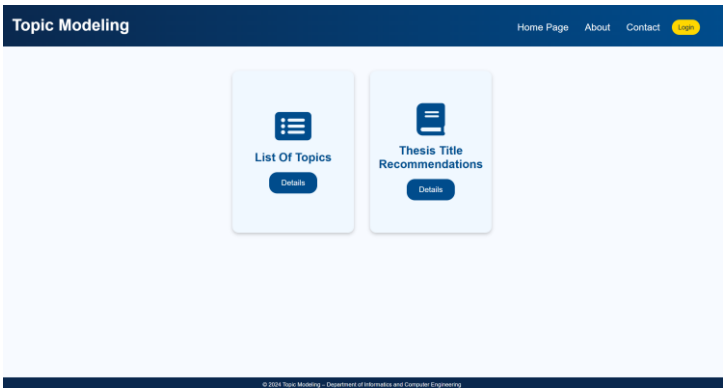
The evaluation results show an agreement percentage of 76.47%, indicating that the majority of topic labels have been assessed as appropriate and relevant by both experts. Thus, the resulting topic labels can be declared qualitatively feasible [26]. However, this study has limitations due to the use of short text data (thesis titles), which have limited context and low co-occurrence rates between words, potentially affecting the LDA model's ability to capture semantic relationships. Consequently, there remains a potential for overlap between topics, and the results must be interpreted with caution due to the model's sensitivity to parameters.

3.3. System development results

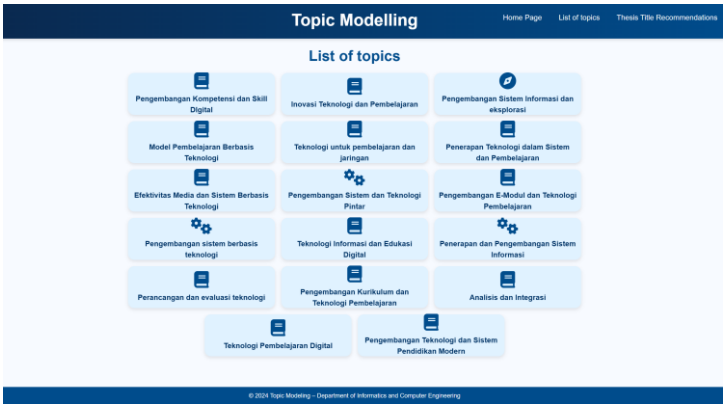
After the tagging process, a web-based system was developed that is integrated with the LDA model so that the research results are applicable and support academic decision-making in exploring and formulating thesis topics. The product, in the form of a website for searching student thesis topics, represents the practical application of topic modeling in higher education, with the interface, navigation, and main functions shown in Figure 5.

Figure 5 shows the interface and main features of the website as a medium for exploring thesis topics and titles. Figure 5(a) shows the main page with two menus, namely topic list and title recommendations, as the initial navigation of the system. Figure 5(b) shows the topic list page, which contains LDA model topic categories and a list of related thesis titles. Figure 5(c) shows the title list page, which presents thesis titles according to the selected topic in a table containing the number, author, year, title, and PDF link sourced from JTIK E-Scription, and is equipped with a search feature. Furthermore, Figure 5(d) shows the title recommendations page, which predicts topics based on the titles entered by users, displaying topic probabilities and a list of relevant titles as research references. After the development stage was completed, system testing was conducted to evaluate the website's ease of use and usefulness through the system usability scale (SUS) method, involving users directly. The statement instruments for the SUS are as shown in Table 11.

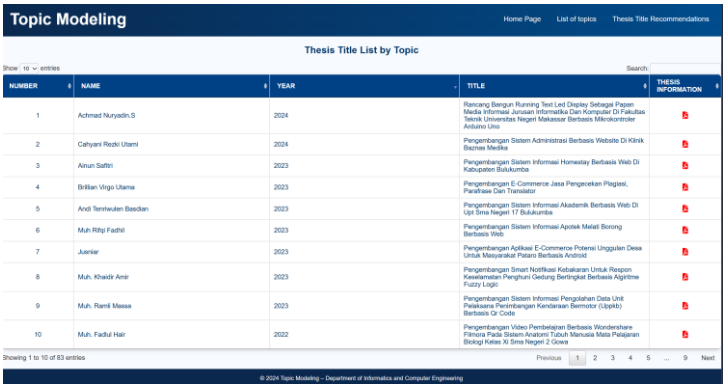
Based on the SUS instrument, respondents were asked to provide ratings using a scale of 1–5 according to their level of agreement with each statement related to the website for exploring student thesis topics that was developed, with the research object being the website itself and a total of 10 respondents. The results of the SUS score calculations for each respondent are presented in Table 12. Based on the results of the SUS score calculation, the system obtained an average score of 77, which falls into the B grade category and is at an acceptable level [27]. These results indicate that the system has a good level of usability and is acceptable to users.



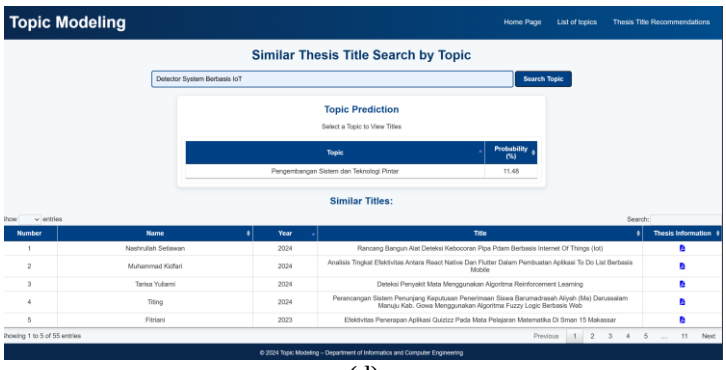
(a)



(b)



(c)



(d)

Figure 5. Display of; (a) the main page, (b) list of topics, (c) list of titles for topics, and (d) title predictions

Table 11. SUS statement instrument

No	Statement	Score
1	I want to use this website often.	1-5
2	I find this website too complicated to use.	1-5
3	I find this website easy to use.	1-5
4	I need help from others to be able to use this website.	1-5
5	I feel that the features on this website are well integrated.	1-5
6	I feel there are too many inconsistencies in this website.	1-5
7	I feel that most people will be able to learn this website quickly.	1-5
8	I feel this website is confusing to use.	1-5
9	I feel confident when using this website.	1-5
10	I need to learn a lot of things first before I can use this website.	1-5

Table 12. SUS test results

Respondents	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score = (total×2.5)
R1	4	2	4	2	4	2	4	4	4	3	82.5
R2	2	3	3	3	3	3	3	3	3	4	75
R3	4	2	5	4	3	3	4	2	3	3	82.5
R4	5	1	4	2	4	1	5	1	4	2	72.5
R5	2	4	4	3	4	2	4	1	4	2	75
R6	4	1	5	1	5	3	5	1	4	1	75
R7	5	1	5	1	4	2	5	1	5	1	75
R8	4	1	5	1	5	1	5	1	4	3	75
R9	4	2	5	1	5	2	5	1	5	2	80
R10	3	2	4	2	4	2	5	2	4	3	77.5
Total											770
Average SUS score											77

4. CONCLUSION

The results of the LDA method implementation test produced the highest coherence value of 0.7431, which indicates a fairly good level of interrelation between words in each identified topic. This coherence value was obtained using a high α parameter of 0.81. This high α parameter supports a more even distribution of topics across documents, creating a balance in topic allocation between documents. On the other hand, a low β parameter (0.01) contributes to topic specificity, resulting in more focused dominant words in each topic. This helps clarify and strengthen the identification of more specific and focused topics. The optimal number of topics 17, was found to create a balance between topics. Additionally, this study provides scientific novelty through a comprehensive parameter optimization process on a large-scale thesis title dataset, resulting in higher coherence compared to previous studies, while also strengthening the effectiveness of LDA in mapping academic research trends more accurately.

FUNDING INFORMATION

The authors declare no funding was involved.

AUTHOR CONTRIBUTION STATEMENT

This journal uses the Contributor Role Taxonomy (CRediT) to acknowledge individual author contributions, reduce membership disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ruslan	✓					✓		✓	✓				✓	✓
Rezki Nurul Jariah S. Intam		✓	✓			✓	✓	✓	✓	✓	✓			
Sasmita	✓		✓	✓			✓	✓		✓	✓		✓	
Dewi Fatmarani Surianto	✓	✓		✓	✓	✓				✓		✓		✓
Andi Akram Nur Risal	✓	✓		✓	✓		✓		✓			✓		✓
Nur Azizah Eka Budiarti	✓	✓		✓		✓			✓	✓			✓	

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The author declares that there is no conflict of interest.

DATA AVAILABILITY

The data supporting this study are available from the corresponding author upon reasonable request. Due to institutional regulations, the dataset is not publicly available and may only be accessed with appropriate permission. All data were collected under official university research authorization in compliance with ethical and confidentiality requirements.

REFERENCES

[1] F. R. B. Arafah *et al*., "Building optimism process in final-year students to finish their theses: case study of Hasanuddin University students threatened with dropout," *International Conference on Medical Imaging, Electronic Imaging, Information Technologies, and Sensors (MIEITS 2024). SPIE*, p. 39, Jun. 26, 2024. doi: 10.1117/12.3030876.

[2] S. Dupont, B. Galand, and F. Nils, "The impact of different sources of social support on academic performance: Intervening factors and mediated pathways in the case of master's thesis," *European Review of Applied Psychology*, vol. 65, no. 5, pp. 227–237, Sep. 2015, doi: 10.1016/j.erap.2015.08.003.

[3] A. Ramírez-Adrados, V. E. Fernández-Elías, S. Fernández-Martínez, B. Martínez-Pascual, C. Gonzalez-de-Ramos, and V. J. Clemente-Suárez, "The effect of studying a double degree in the psychophysiological stress response in the bachelor's thesis defense," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, p. 1207, Jan. 2022, doi: 10.3390/ijerph19031207.

[4] C. T. Suryawati, R. Purwaningrum, and A. T. Susilo, "Developing anxiety scale of writing thesis on students," *KONSELI : Jurnal Bimbingan dan Konseling (E-Journal)*, vol. 7, no. 2, pp. 163–170, Nov. 2020, doi: 10.24042/kons.v7i2.7578.

[5] R. K. Saini *et al*., "Selecting a thesis topic: a postgraduate's dilemma," *Industrial Psychiatry Journal*, vol. 32, no. 1, pp. 19–23, Jan. 2023, doi: 10.4103/ipj.ipj_62_22.

[6] Z. I. Muslimin, P. Pihasniwati, E. L. Widuri, and B. J. Anilon, "Positive thinking as a mediator in the relationship faith and resilience in students who are working on a thesis," *International Journal of Islamic Educational Psychology*, vol. 6, no. 1, pp. 137–153, Jul. 2025, doi: 10.18196/ijiep.v6i1.25059.

[7] D. Nasien and E. Putra Yansen, "Implementation of the winnowing algorithm in detecting plagiarism in title and abstract of student's final project," *IT Journal Research and Development*, vol. 7, no. 1, pp. 23–35, Aug. 2022, doi: 10.25299/itjrd.2022.7897.

[8] M. Kokoç, M. Kokoç, and Ö. Tuncer, "Using Text Mining to Identify Research Trends in Management Information Systems Theses: A Topic Modeling Approach," *Ordu Üniversitesi Bilim ve Teknoloji Dergisi*, vol. 14, no. 2, pp. 293–306, Dec. 2024, doi: 10.54370/ordubtd.1598387.

[9] N. S. M. Nik Mangsor, S. A. Md Nasir, W. F. Wan Yaacob, Z. Ismail, and S. Abdul Rahman, "Analysing corporate social responsibility reports using document clustering and topic modeling techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 3, p. 1546, Jun. 2022, doi: 10.11591/ijeecs.v26.i3.pp1546-1555.

[10] A. Goyal and I. Kashyap, "A data-driven analysis to determine the optimal number of topics 'K' for latent Dirichlet allocation model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 1, p. 310, Jul. 2024, doi: 10.11591/ijeecs.v35.i1.pp310-322.

[11] V. Gangadharan and D. Gupta, "Recognizing named entities in agriculture documents using LDA based topic modelling techniques," *Procedia Computer Science*, vol. 171, pp. 1337–1345, 2020, doi: 10.1016/j.procs.2020.04.143.

[12] D. Pruss *et al*., "Zika discourse in the Americas: A multilingual topic analysis of Twitter," *PLoS ONE*, vol. 14, no. 5, p. e0216922, May 2019, doi: 10.1371/journal.pone.0216922.

[13] K. Ylikruuvi, K. Järvelin, P. Vakkari, and M. Juhola, "Comparing representations of a discipline derived through LDA vs. intellectual content analysis: the case of information science," *Scientometrics*, vol. 130, no. 8, pp. 4309–4337, Jul. 2025, doi: 10.1007/s11192-025-05376-1.

[14] R. K. Gupta, R. Agarwalla, B. H. Naik, J. R. Evuri, A. Thapa, and T. D. Singh, "Prediction of research trends using LDA based topic modeling," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 298–304, Jun. 2022, doi: 10.1016/j.gltp.2022.03.015.

[15] M. Masinde, "Application of Latent Dirichlet Allocation Topic Model in Identifying 4IR Research Trends," *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Nature Switzerland*, pp. 74–94, 2023. doi: 10.1007/978-3-031-35883-8_6.

[16] M. Mujahid *et al*., "Sentiment analysis and topic modeling on tweets about online education during COVID-19," *Applied Sciences*, vol. 11, no. 18, p. 8438, Sep. 2021, doi: 10.3390/app11188438.

[17] E. Christodoulou, A. Gregoriades, M. Pampaka, and H. Herodotou, "Combination of topic modelling and decision tree classification for tourist destination marketing," in *Lecture Notes in Business Information Processing*, vol. 382 LNBIP, 2020, pp. 95–108, doi: 10.1007/978-3-030-49165-9_9.

[18] N. Z. Dina, S. D. Ravana, and N. Idris, "Legal Judgment Prediction using Natural Language Processing and Machine Learning Methods: A Systematic Literature Review," *Sage Open*, vol. 15, no. 2, Apr. 2025, doi: 10.1177/21582440251329663.

[19] I. M. K. Karo, M. F. M. Fudzee, S. Kasim, and A. A. Ramli, "Karonese sentiment analysis: a new dataset and preliminary result," *International Journal on Informatics Visualization*, vol. 6, no. 2–2, pp. 523–530, 2022, doi: 10.30630/joiv.6.2-2.1119.

[20] Z. Yu, T. R. Johnson, and R. Kavuluru, "Phrase based topic modeling for semantic information processing in biomedicine," in *Proceedings - 2013 12th International Conference on Machine Learning and Applications, ICMLA 2013*, Dec. 2013, vol. 1, pp. 440–445, doi: 10.1109/ICMLA.2013.89.
[21] Y. Tekin, "Optimization of LDA parameters," in *2020 28th Signal Processing and Communications Applications Conference (SIU)*, Oct. 2020, pp. 1–4, doi: 10.1109/SIU49456.2020.9302034.
[22] J. Gan and Y. Qi, "Selection of the optimal number of topics for LDA topic model—Taking patent policy analysis as an example," *Entropy*, vol. 23, no. 10, p. 1301, Oct. 2021, doi: 10.3390/e23101301.
[23] M. S. Kaiser, A. Bandyopadhyay, M. Mahmud, and K. Ray, "Proceedings of International Conference on Trends in Computational and Cognitive Engineering," *Proceedings of TCCE 2020, Springer Singapore*, 2021. doi: 10.1007/978-981-33-4673-4.
[24] D. M. Blei and J. D. Lafferty, "A correlated topic model of Science," *The Annals of Applied Statistics*, vol. 1, no. 1, Jun. 2007, doi: 10.1214/07-AOAS114.
[25] K. Tran, "pyLDAvis: Topic Modelling Exploration Tool That Every NLP Data Scientist Should Know," *Neptune.Ai*, 2023. https://neptune.ai/blog/pyldavis-topic-modelling-exploration-tool-that-every-nlp-data-scientist-should-know (Accessed: 04 February, 2025).
[26] A. M. Marchevsky, A. E. Walts, B. I. Lissenberg-Witte, and E. Thunnissen, "Pathologists should probably forget about kappa. Percent agreement, diagnostic specificity and related metrics provide more clinically applicable measures of interobserver variability," *Annals of Diagnostic Pathology*, vol. 47, p. 151561, Aug. 2020, doi: 10.1016/j.anndiagpath.2020.151561.
[27] F. Olivia and A. Ibrahim, "Evaluating user experience and usability of the USEPT website using user experience questionnaire and system usability scale method," *Journal of Information Systems and Informatics*, vol. 6, no. 4, pp. 2632–2648, Dec. 2024, doi: 10.51519/journalisi.v6i4.910.

BIOGRAPHIES OF AUTHORS

Ruslan obtained his Bachelor's, Master's, and Doctorate degrees in Electrical Engineering Education from Makassar State University. Currently, he serves as a Professor and Lecturer in the Department of Electrical Engineering Education, Faculty of Engineering, Makassar State University. Additionally, he serves as Dean of the Faculty of Teacher Education and Training (FKIP) at the University of West Sulawesi (UNSULBAR), Chair of the UNSULBAR Senate, and Vice Rector II of UNSULBAR. His research interests include educational technology, education, and power electronics. He can be contacted at email: ruslan.ft@unm.ac.id.

Rezki Nurul Jariah S. Intam obtained a Bachelor's degree in Computer Engineering from the University of Makassar. Her research interests include natural language processing (NLP). She can be contacted at email: rezkinuruljariah@gmail.com.

Sasmita obtained a Bachelor's degree in Computer Engineering from Makassar State University. Her research interests include data science and data analysis. She can be contacted at email: sasmita.raca@gmail.com.

Dewi Fatmarani Surianto is a lecturer in the Computer Engineering Study Program, Department of Informatics and Computer Engineering, Faculty of Engineering, Makassar State University. She obtained her Bachelor's degree in Informatics from Bakrie University in 2016, after beginning her studies in 2012. She then completed her Master's degree in Informatics at Telkom University, graduating in 2019. She began her academic career in 2021 as a lecturer at Kalla Institute, a private higher education institution. Since joining Makassar State University, she has been actively involved in various research and community service initiatives. Her current research interests primarily focus on the development of artificial intelligence (AI), with a particular emphasis on natural language processing (NLP) and data science. She can be contacted at email: dewifatmaranis@unm.ac.id.

Andi Akram Nur Risal obtained his Master's degree in Computer Science. He currently serves as a Lecturer in the Computer Engineering Study Program at Makassar State University. His research interests include data mining, algorithm development, natural language processing (NLP), and machine learning. He can be contacted at email: akramandi@unm.ac.id.

Nur Azizah Eka Budiarti obtained a Bachelor's degree in Informatics Engineering from the State Islamic University (UIN) Alauddin Makassar in 2018 and a Master's degree in Computer Science from IPB University in 2022. Currently, she serves as a Lecturer in the Computer Engineering Study Program, Faculty of Engineering, Makassar State University. Her research interests include artificial intelligence, expert systems, and computer vision. She can be contacted at email: nurazizah.ekabudiarti@gmail.com.