

TOPIC ANALYSIS OF STUDENT FEEDBACK ON LEARNING MANAGEMENT SYSTEMS USING BERTOPIC: A COMPARATIVE STUDY OF INDOBERT, DISTILBERT, AND SBERT

Aldi Aditya Perdana¹, Sajarwo Anggai², Winarni³

^{1,2,3}Universitas Pamulang, Tangerang Selatan, Indonesia

Email: aldiadityaperdana97@gmail.com

Abstract

The widespread adoption of Learning Management Systems (LMS) in digital education has generated large volumes of student feedback in the form of unstructured free-text data, making manual analysis increasingly impractical. This study aims to identify the dominant themes emerging from student feedback on LMS platforms and to compare the performance of different Transformer-based embedding models in topic modeling tasks. The proposed approach employs BERTopic with three embedding models, namely IndoBERT, DistilBERT, and Sentence-BERT (SBERT). Student feedback data were collected from an institutional LMS and processed through text preprocessing, embedding generation, and topic modeling stages. Model performance was evaluated using multiple coherence metrics (c_v , c_{npmi} , u_{mass} , and c_{uci}), topic diversity, and the proportion of outlier documents. The results indicate that the IndoBERT-family model achieves the highest coherence scores, particularly in c_v and c_{npmi} , suggesting superior semantic consistency of the generated topics. DistilBERT produces the lowest proportion of outliers but yields a more limited number of topics, while SBERT demonstrates a balanced performance between topic quality and thematic diversity. These findings highlight that the choice of embedding model significantly influences the quality of topic modeling outcomes for Indonesian-language student feedback data.

Keywords: *Topic Modeling, BERTopic, Learning Management System, Transformer Models, Student Feedback.*

A. INTRODUCTION

The rapid advancement of information technology has accelerated the widespread adoption of Learning Management Systems (LMS) in educational environments as a core component of digital learning. LMS platforms are not only utilized for content distribution and learning activity management but also serve as a medium for students to express feedback regarding their learning experiences. Such feedback plays a crucial role in supporting educators and educational institutions in evaluating instructional materials, teaching methods, and overall learning effectiveness.

However, student feedback collected through LMS platforms is predominantly expressed in the form of free-text responses that are unstructured, short, and highly diverse in both length and context. These characteristics make manual analysis time-consuming, subjective, and increasingly impractical, particularly when dealing with large-scale datasets. Consequently, automated text analysis techniques are required to

efficiently extract meaningful patterns and dominant themes from student feedback in an objective and scalable manner.

Topic modelling is a well-established approach in Natural Language Processing (NLP) for uncovering latent thematic structures within large collections of text documents. Traditional probabilistic methods, such as Latent Dirichlet Allocation (LDA), have been widely applied in various domains. Nevertheless, these methods exhibit limitations in capturing contextual and semantic relationships, especially when applied to short and informal texts, which are common in student feedback data.

Recent developments in NLP have introduced Transformer-based language models, such as BERT and its variants, which offer context-aware semantic representations and significantly improve text understanding. Building upon these advances, BERTopic has emerged as a modern topic modelling framework that integrates Transformer-based embeddings with dimensionality reduction and clustering techniques to generate more semantically coherent topics. BERTopic has demonstrated promising performance across multiple domains, particularly in handling short-text corpora.

In the context of Indonesian-language text analysis, the selection of embedding models is a critical factor due to linguistic characteristics that differ from English. IndoBERT is specifically pre-trained on Indonesian corpora, enabling it to capture language-specific semantics more effectively. In contrast, DistilBERT offers a computationally efficient alternative with reduced model complexity, while Sentence-BERT (SBERT) is designed to generate robust sentence-level embeddings suitable for semantic similarity tasks.

Based on these considerations, this study focuses on analyzing student feedback on LMS platforms using BERTopic by conducting a comparative evaluation of three Transformer-based embedding models: IndoBERT, DistilBERT, and SBERT. The primary objective is to investigate how different embedding models influence topic coherence, thematic diversity, and clustering stability. By providing a systematic comparison, this research aims to contribute to the application of Transformer-based topic modelling in Indonesian educational data analysis and to support the development of more effective data-driven evaluation systems for digital learning environments.

B. METHOD

The research was conducted through several interconnected stages, ranging from data collection to the evaluation of topic modeling results. The overall research workflow is illustrated in Figure 1.

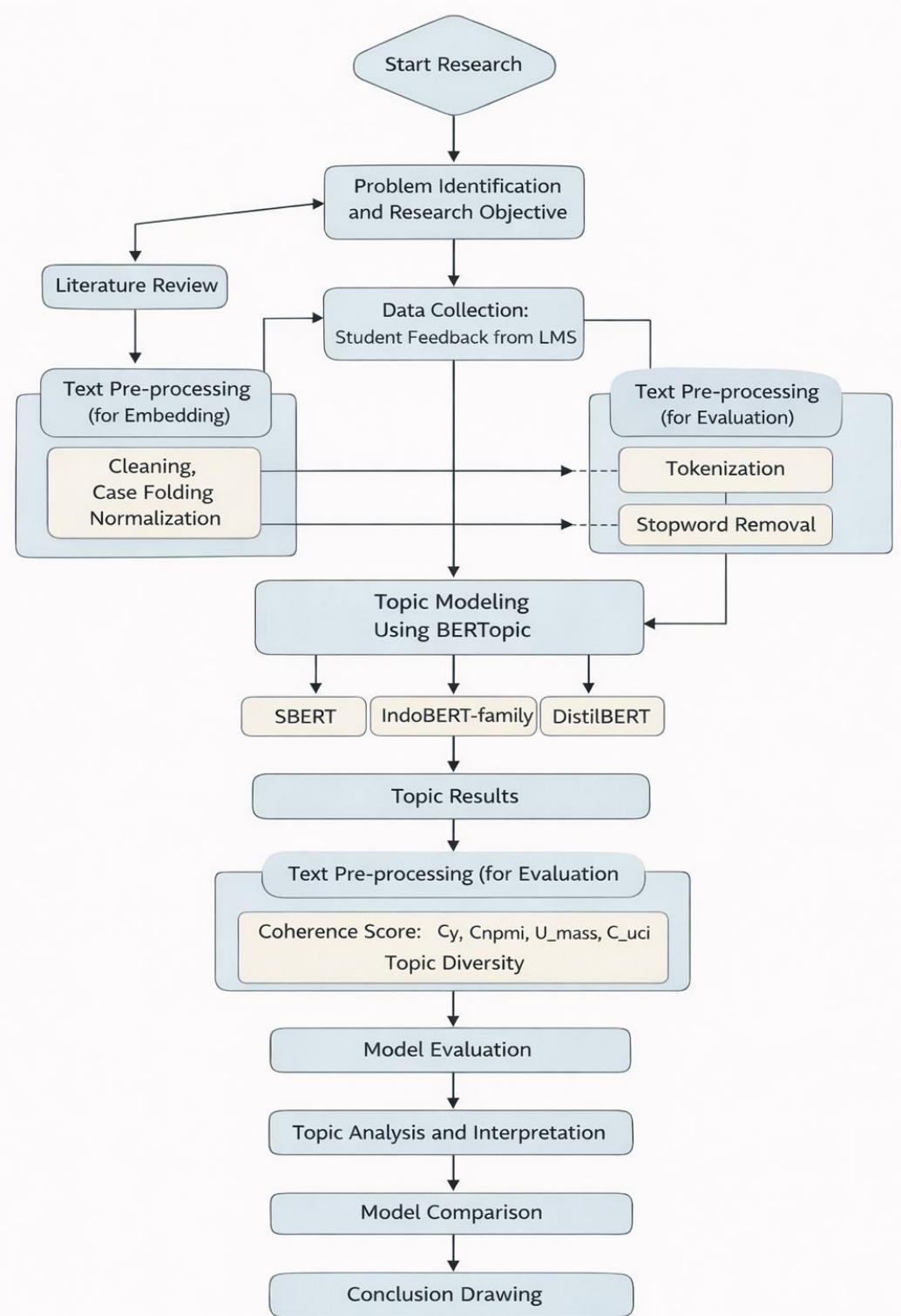


Figure 1. Research Workflow for Analyzing Student Feedback Topics Using BERTopic

As shown in Figure 1, the research process begins with the collection of student feedback text data obtained from a Learning Management System (LMS). The collected data then undergo a preprocessing stage aimed at improving data quality. Subsequently, the pre-processed texts are transformed into vector representations

using several Transformer-based embedding models. These vector representations are further processed using the BERTopic method to generate the main topics. The resulting topics are evaluated using coherence metrics, topic diversity, and outlier analysis to compare the performance of each embedding model.

This study adopts a quantitative approach with an experimental method to analyze and compare the performance of Transformer-based embedding models in topic modelling tasks. The primary focus of the research is the application of BERTopic to extract dominant themes from student feedback related to the Learning Management System (LMS).

The data used in this study consist of a collection of student feedback texts obtained from an LMS within an educational environment. The feedback data are generally short texts that reflect students' experiences, opinions, and evaluations of the learning process. All data were analyzed anonymously and in aggregated form to ensure the confidentiality of the respondents.

The initial stage of the research involves text preprocessing to enhance data quality prior to analysis. This stage includes the removal of non-alphabetic characters, text normalization, stop-word removal, and text formatting adjustments to ensure compatibility with the embedding process. These steps aim to reduce noise and produce more consistent text representations.

Next, the pre-processed student feedback texts are converted into vector representations using three Transformer-based embedding models: IndoBERT-family, DistilBERT, and Sentence-BERT (SBERT). IndoBERT-family is used to capture Indonesian-specific semantic contexts, DistilBERT is selected for its computational efficiency, and SBERT is employed for its ability to generate stable and representative sentence embeddings. The embedding process is conducted separately for each model to allow for an objective comparison of topic modelling results.

The resulting vector representations are then processed using BERTopic to generate the main topics. BERTopic integrates dimensionality reduction through Uniform Manifold Approximation and Projection (UMAP) with clustering using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). This approach enables document grouping based on semantic similarity while identifying documents that do not belong to any cluster as outliers.

Model performance is evaluated using several quantitative metrics. Topic quality is measured using coherence scores, including c_v , c_{npmi} , u_{mass} , and c_{uci} , to assess semantic consistency among words within each topic. In addition, topic diversity is used to measure the variation of keywords across topics, while the proportion of outlier documents is analyzed to evaluate clustering stability. These metrics serve as the basis for comparing the performance of the embedding models.

To support result analysis and interpretation, this study also develops a visualization dashboard based on Flask that presents summaries of evaluation metrics and comparative graphs across models. The dashboard functions solely as a result exploration tool and is not used for model training, thereby maintaining a clear separation between analysis and presentation stages.

C. RESULT AND DISCUSSION

This section presents the results of topic modeling performed on student feedback data collected from the Learning Management System (LMS) using the BERTopic framework with three Transformer-based embedding models: IndoBERT-family, DistilBERT, and Sentence-BERT (SBERT). The evaluation focuses on topic coherence, topic diversity, and the proportion of outlier documents to assess the quality, diversity, and stability of the generated topics.

Topic Coherence Evaluation

The coherence score evaluation indicates that the IndoBERT-family model achieves the highest performance in terms of semantic coherence, particularly on the c_v and c_{npmi} metrics. Higher c_v values suggest that the words within each generated topic are more semantically related and interpretable, while higher c_{npmi} values indicate stronger co-occurrence relationships among topic keywords within the dataset.

Table 1. Coherence Score Results for Each Embedding Model

| Model | c_v | c_{npmi} | u_{mass} | c_{uci} |
|-----------------|--------|------------|------------|-----------|
| DistilBERT | 0,5602 | 0,0079 | -1,1878 | 0,0696 |
| IndoBERT-family | 0,5933 | 0,0531 | -2,7557 | -0,2606 |
| SBERT | 0,5604 | 0,0400 | -2,9781 | -0,5936 |

DistilBERT and SBERT exhibit relatively lower coherence scores on these metrics. Nevertheless, DistilBERT maintains a competitive c_v value, indicating that model compression does not significantly degrade semantic representation quality. In contrast, all models produce negative values for the u_{mass} and c_{uci} metrics. Such results are commonly observed when analyzing short and informal text data, such as student feedback, and therefore these metrics are treated as complementary rather than primary indicators.

Topic Diversity Evaluation

Topic diversity is used to measure the degree of lexical variation among topics generated by each embedding model. A higher topic diversity score indicates that the topics contain more distinct keywords and exhibit lower redundancy.

Table 2. Topic Diversity Scores of Embedding Models

| Model | Topic Diversity |
|-----------------|-----------------|
| DistilBERT | 0,6333 |
| IndoBERT-family | 0,5269 |
| SBERT | 0,5241 |

In contrast, IndoBERT-family and SBERT yield lower topic diversity scores, implying greater similarity in vocabulary across topics. However, lower diversity does not necessarily reflect poorer topic quality, as it may indicate a more focused thematic structure aligned with the LMS learning context.

Topic Interpretation

Topic interpretation is conducted by analyzing the dominant keywords associated with each topic generated by the model. Based on the evaluation results,

IndoBERT-family is selected as the primary model for in-depth topic analysis due to its superior semantic coherence and relevance to Indonesian-language contexts.

Table 3. Summary of the Main Topics Generated Using IndoBERT-family

| Topic Code | Dominant Keywords | Topic Interpretation | Relevance to LMS Evaluation |
|------------|--------------------------------|------------------------------|--|
| T1 | material, computer, device | Basic ICT learning materials | Alignment of learning content |
| T2 | scratch, game, coding | Practice-based learning | Interactive learning methods |
| T3 | word, excel, power point | Office application skills | Supporting student competencies |
| T4 | logic, algorithm, fundamentals | Basic programming concepts | Enhancement of logical thinking skills |

Comparison of Embedding Models

Overall, each embedding model exhibits distinct characteristics in topic modeling performance. IndoBERT-family produces topics with the highest semantic coherence and contextual relevance but shows a relatively higher proportion of outlier documents. DistilBERT demonstrates greater clustering stability, as indicated by the lowest outlier proportion, although it generates fewer topics. SBERT offers a balanced performance, achieving moderate coherence, topic diversity, and outlier proportions.

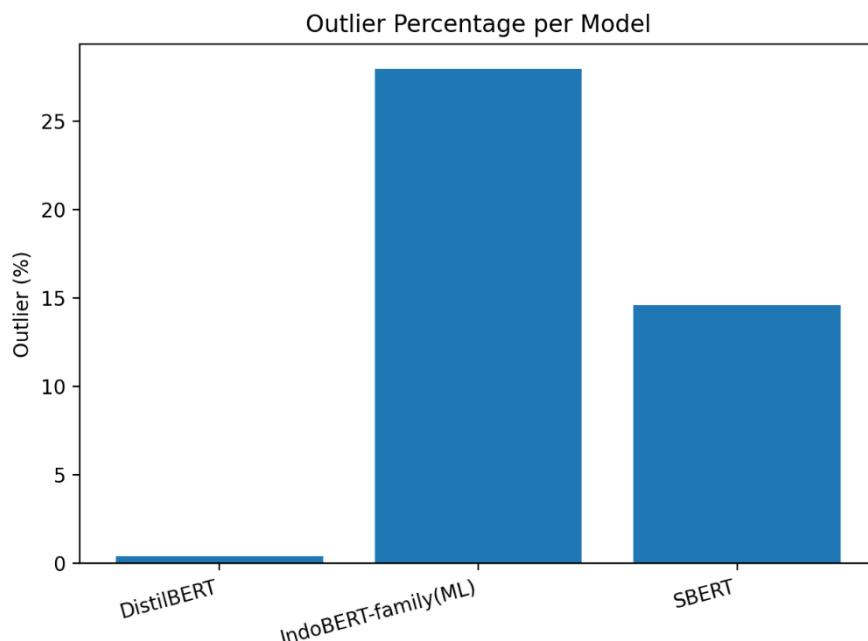


Figure 2. Percentage of Outlier Documents Produced by Each Embedding Model

Based on the research objectives, IndoBERT-family is identified as the most suitable embedding model for analyzing Indonesian-language student feedback data, as it provides semantically coherent and interpretable topics relevant to LMS evaluation.

D. CONCLUSION

This study analyzed the dominant themes present in student feedback on a Learning Management System (LMS) and compared the performance of several Transformer-based embedding models within the BERTopic framework. The results demonstrate that BERTopic is effective in extracting meaningful and interpretable topics from short and informal student feedback texts.

The evaluation results indicate that the IndoBERT-family model produces the most semantically coherent topics, particularly based on the c_v and c_{npmi} coherence metrics. This finding highlights the advantage of using language-specific embedding models when analyzing Indonesian-language educational data. The topics generated by IndoBERT-family are more contextually relevant and easier to interpret in relation to LMS-based learning evaluation.

DistilBERT shows strengths in clustering stability, as reflected by the lowest proportion of outlier documents and the highest topic diversity score. These results suggest that DistilBERT is capable of generating diverse topics with minimal redundancy, although the number of topics produced is relatively limited. Meanwhile, SBERT demonstrates balanced performance across coherence, diversity, and outlier metrics, positioning it as a flexible alternative for topic modelling tasks.

Overall, the findings confirm that the choice of embedding model significantly influences the quality and characteristics of topic modelling results. For the analysis of Indonesian-language student feedback on LMS platforms, the IndoBERT-family model is the most suitable option due to its superior semantic coherence and contextual relevance. This study provides valuable insights for the development of text-based learning evaluation systems and may serve as a reference for future research on topic modelling in educational contexts.

REFERENCES

Abdurrazzaq, M. A. (n.d.). Analisis Ulasan Aplikasi MyPertamina Menggunakan *Topic modeling* dengan Latent Dirichlet Allocation. *Jurnal Sains Dan Teknologi*, 10(1).

Abella, Á. R., Silvestre, J. P., & Tabuada, P. (2024). *The Asymptotic Behavior of Attention in Transformers*. 1–26. <http://arxiv.org/abs/2412.02682>

Ahammad, T. (2024). Identifying hidden patterns of fake COVID-19 news: An in-depth sentiment analysis and *topic modeling* approach. *Natural Language Processing Journal*, 6(January), 100053. <https://doi.org/10.1016/j.nlp.2024.100053>

Akdeas Oktanae Widodo, Septiadi, F., & Nur Aini Rakhmawati. (2023). Analisis Tren Konten Pada Vtuber Indonesia Menggunakan Latent Dirichlet Allocation. *Jurnal Informatika Dan Rekayasa Elektronik*, 6(1), 56–63. <https://doi.org/10.36595/jire.v6i1.718>

Alamsyah, A., & Girawan, N. D. (2023). Improving Clothing Product Quality and Reducing Waste Based on Consumer Review Using RoBERTa and BERTopic Language Model. *Big Data and Cognitive Computing*, 7(4). <https://doi.org/10.3390/bdcc7040168>

Allenbrand, C. (2024). Supervised and unsupervised learning models for pharmaceutical drug rating and classification using consumer generated

ARTIKEL

reviews. *Healthcare Analytics*, 5(December 2023), 100288. <https://doi.org/10.1016/j.health.2023.100288>

Alonso-Dos-Santos, M., Sánchez Franco, M., Calabuig, F., & González-Serrano, M. H. (2023). Modelling the structure of the sports management research field using the bertopic approach. *Retos*, 47, 648–663. <https://doi.org/10.47197/retos.v47.93622>

Alryalat, S. A., Qasem, A., Albdour, K., & Rawashdeh, B. (2023). Assessment of Topics Published in Leading Medical Journals Using Natural Language Processing. *High Yield Medical Reviews*, 1(1), 1–8. <https://doi.org/10.59707/hymrhmdo2739>

An, Y., Oh, H., & Lee, J. (2023). Marketing Insights from Reviews Using Topic modeling with BERTopic and Deep Clustering Network. *Applied Sciences (Switzerland)*, 13(16). <https://doi.org/10.3390/app13169443>

Arif Fitra Setyawan, Amelia Devi Putri Ariyanto, Fari Katul Fikriah, & Rozaq Isnaini Nugraha. (2024). Analisis Sentimen Ulasan iPhone di Amazon Menggunakan Model Deep Learning BERT Berbasis Transformer. *Elkom: Jurnal Elektronika Dan Komputer*, 17(2), 447–452. <https://doi.org/10.51903/elkom.v17i2.2150>

Arslan, M., & Cruz, C. (2023). Leveraging NLP approaches to define and implement text relevance hierarchy framework for business news classification. *Procedia Computer Science*, 225, 317–326. <https://doi.org/10.1016/j.procs.2023.10.016>

Aryadi, J. A., Basith, Y. A. A., Munawir, M., & Agustini, D. A. R. (2023). Analisis Data Review Hotel di Google Maps Melalui Text Mining (Studi Kasus: Kabupaten Bandung). *JIKO (Jurnal Informatika dan Komputer)*, 7(2), 312. <https://doi.org/10.26798/jiko.v7i2.938>

Bachoumis, A., Mylonas, C., Plakas, K., Birbas, M., & Birbas, A. (2023). Data-Driven Analytics for Reliability in the Buildings-to-Grid Integrated System Framework: A Systematic Text-Mining-Assisted Literature Review and Trend Analysis. *IEEE Access*, 11(October), 130763–130787. <https://doi.org/10.1109/ACCESS.2023.3335191>

Bu, W., Shu, H., Kang, F., Hu, Q., & Zhao, Y. (2023). Software Subclassification Based on BERTopic-BERT-BiLSTM Model. *Electronics (Switzerland)*, 12(18). <https://doi.org/10.3390/electronics12183798>

Chagnon, E., Pandolfi, R., Donatelli, J., & Ushizima, D. (2024). Benchmarking topic models on scientific articles using BERTeley. *Natural Language Processing Journal*, 6(October 2023), 100044. <https://doi.org/10.1016/j.nlp.2023.100044>

Colantoni, F. (2023). The impact of corporate governance on default risk: BERTopic literature review. *Corporate Ownership and Control*, 20(4), 57–71. <https://doi.org/10.22495/cocv20i4art4>

Dahlke, J. (2024). *Artificial intelligence as a sociotechnical system: Integrating technical design, human goals, and social expectations*. Journal of Artificial Intelligence Studies, 12(1), 45–60.

Herwinskyah. (2023). Pemodelan Topik Dalam Al-Qur'an Menggunakan Library. *Simetris*, 14(2), 319–327.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. arXiv preprint arXiv:2203.05794. <https://arxiv.org/abs/2203.05794>

ARTIKEL

Jamaaluddin, & Sulistyowati, I. (2021). Buku Ajar Kecerdasan Buatan. *Umsida Press*, 121.

Khadijah, U. N., & Cahyono, N. (2024). Analisis Topic Modelling Pariwisata Yogyakarta Menggunakan Latent Dirichlet Allocation (LDA). *The Indonesian Journal of Computer Science*, 13(4).

Masruriyah, A. F. N., Sukmawati, C. E., & Novita, H. Y. (2022). Pengelompokan Topik Cuitan Pengguna Twitter Terhadap Kuliah Kerja Nyata (KKN) di Indonesia Menggunakan Latent Dirichlet Allocation. *Konferensi Nasional Penelitian Dan Pengabdian (KNPP)*, 3, 1128–1133.

Matira, Y., & Setiawan, I. (2023). Pemodelan Topik pada Judul Berita Online Detikcom Menggunakan Latent Dirichlet Allocation. *Estimasi: Journal of Statistics and Its Application*, 4(1), 2379–2721. <https://doi.org/10.20956/ejsa.vi.24843>

Maulidiya, D. (2023). Topic Modelling using Latent Dirichlet Allocation (LDA) to Investigate the Latent Topics of Mathematical Creative Thinking Research in Indonesia. *Journal of Intelligent Computing & Health Informatics*, 3(2), 35. <https://doi.org/10.26714/jichi.v3i2.11428>

Mueller, J. P., & Massaron, L. (2018). *Artificial Intelligence for Dummies*. John Wiley & Sons, Inc.

Mulia, A., & Dzikrillah, A. R. (2023a). Analisis Perbedaan Pendapat Netizen Indonesia tentang Presiden Jokowi sebelum dan sesudah Kenaikan Harga BBM Analysis of Indonesian Netizens' Dissent on President Jokowi before and after Fuel Price Increase. *Journal of Computing Engineering, System and Science*, 8(2), 318–328.

Mulia, A., & Dzikrillah, A. R. (2023b). Analisis Perbedaan Pendapat Netizen Indonesia tentang Presiden Jokowi sebelum dan sesudah Kenaikan Harga BBM Analysis of Indonesian Netizens' Dissent on President Jokowi before and after Fuel Price Increase. *Journal of Computing Engineering, System and Science*, 8(2), 318–328.

Naghshzhan, A., & Ratte, S. (2023). *Enhancing API Documentation through BERTopic modeling and Summarization*. <http://arxiv.org/abs/2308.09070>

Niroomand, K., Saady, N. M. C., Bazan, C., Zendehboudi, S., Soares, A., & Albayati, T. M. (2023). Smart investigation of artificial intelligence in renewable energy system technologies by natural language processing: Insightful pattern for decision-makers. *Engineering Applications of Artificial Intelligence*, 126(PA), 106848. <https://doi.org/10.1016/j.engappai.2023.106848>

Ojo, A. O., & Bouguila, N. (2024). A topic modeling and image classification framework: The Generalized Dirichlet variational autoencoder. *Pattern Recognition*, 146(October 2023), 110037. <https://doi.org/10.1016/j.patcog.2023.110037>

Parlina, A., & Maryati, I. (2023). Leveraging BERTopic for the Analysis of Scientific Papers on Seaweed. *Proceedings - 2023 10th International Conference on Computer, Control, Informatics and Its Applications: Exploring the Power of Data: Leveraging Information to Drive Digital Innovation, IC3INA 2023*, 2022, 279–283. <https://doi.org/10.1109/IC3INA60834.2023.10285737>

Rahman, R. D., Setiawan, N. Y., & Bachtiar, F. A. (2025). Analisis Sentimen Pengguna Aplikasi Mobile Berbasis Review Pada Platform Blibli Menggunakan Metode Bidirectional Encoder Representations from Transformers (BERT). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 9(4), 2548–2964.

ARTIKEL

Rejeb, A., Rejeb, K., Appolloni, A., Jagtap, S., Iranmanesh, M., Alghamdi, S., Alhasawi, Y., & Kayikci, Y. (2024). Unleashing the power of internet of things and blockchain: A comprehensive analysis and future directions. *Internet of Things and Cyber-Physical Systems*, 4(June 2023), 1-18. <https://doi.org/10.1016/j.iotcps.2023.06.003>

Rosalinda, G., Santoso, R., & Kartikasari, P. (2023). Pemodelan Topik Ulasan Aplikasi Netflix Pada Google Play Store Menggunakan Latent Dirichlet Allocation. *Jurnal Gaussian*, 11(4), 554-561. <https://doi.org/10.14710/j.gauss.11.4.554-561>

Saidi, F., Trabelsi, Z., & Thangaraj, E. (2022). A novel framework for semantic classification of cyber terrorist communities on Twitter. *Engineering Applications of Artificial Intelligence*, 115(January), 105271. <https://doi.org/10.1016/j.engappai.2022.105271>

Samsir, S., Saragih, R. S., Subagio, S., Aditiya, R., & Watrianthos, R. (2023). BERTopic modeling of Natural Language Processing Abstracts: Thematic Structure and Trajectory. *Jurnal Media Informatika Budidarma*, 7(3), 1514. <https://doi.org/10.30865/mib.v7i3.6426>

Suryotrisongko, H., Ginardi, H., Ciptaningtyas, H. T., Dehqan, S., & Musashi, Y. (2022). Topic modeling for Cyber Threat Intelligence (CTI). 2022 7th International Conference on Informatics and Computing, ICIC 2022, 1-7. <https://doi.org/10.1109/ICIC56845.2022.10006988>

Tondang, B. A., Fadhil, M. R., Perdana, M. N., Fauzi, A., & Janitra, U. S. (2023). Analisis pemodelan topik ulasan aplikasi BNI, BCA, dan BRI menggunakan latent dirichlet allocation. *INFOTECH: Jurnal Informatika & Teknologi*, 4(1), 114-127. <https://doi.org/10.37373/infotech.v4i1.601>

Wang, Y., Bashar, M. A., Chandramohan, M., & Nayak, R. (2023). Exploring topic models to discern cyber threats on Twitter: A case study on Log4Shell. *Intelligent Systems with Applications*, 20(March), 200280. <https://doi.org/10.1016/j.iswa.2023.200280>

Yvon, F. (2023). Transformers in Natural Language Processing. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13500 LNAI(May), 81-105. https://doi.org/10.1007/978-3-031-24349-3_6

Zain, R. M., Anggai, S., Tukiyat, Musyafa, A., & Waskita, A. A. (2024). Revealing a Country's Government Discourse Through BERT-based Topic modeling in the US Presidential Speeches. *International Conference on Computer, Control, Informatics and Its Applications*, IC3INA, 2024, 191-196. <https://doi.org/10.1109/IC3INA64086.2024.10732578>

Zhang, D., Wu, X., Liu, P., Qin, H., & Zhou, W. (2023). Identification of Product Innovation Path Incorporating the FOS and BERTopic Model from the Perspective of Invalid Patents. *Applied Sciences (Switzerland)*, 13(13). <https://doi.org/10.3390/app13137987>

Zou, T., Guo, P., Li, F., & Wu, Q. (2024). Research topic identification and trend prediction of China's energy policy: A combined LDA-ARIMA approach. *Renewable Energy*, 220(February 2023), 119619. <https://doi.org/10.1016/j.renene.2023.119619>