

**Deteksi Tren Topik Penelitian Bidang Kecerdasan Buatan Menggunakan  
LDA dan BERTopic**

**Laporan Tugas Akhir**

Diajukan Untuk Memenuhi  
Persyaratan Guna Meraih Gelar Sarjana Informatika  
Universitas Muhammadiyah Malang



Rahajeng Febri Shafiyah

202110370311308

**Bidang minat :**

Data Science

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MALANG**

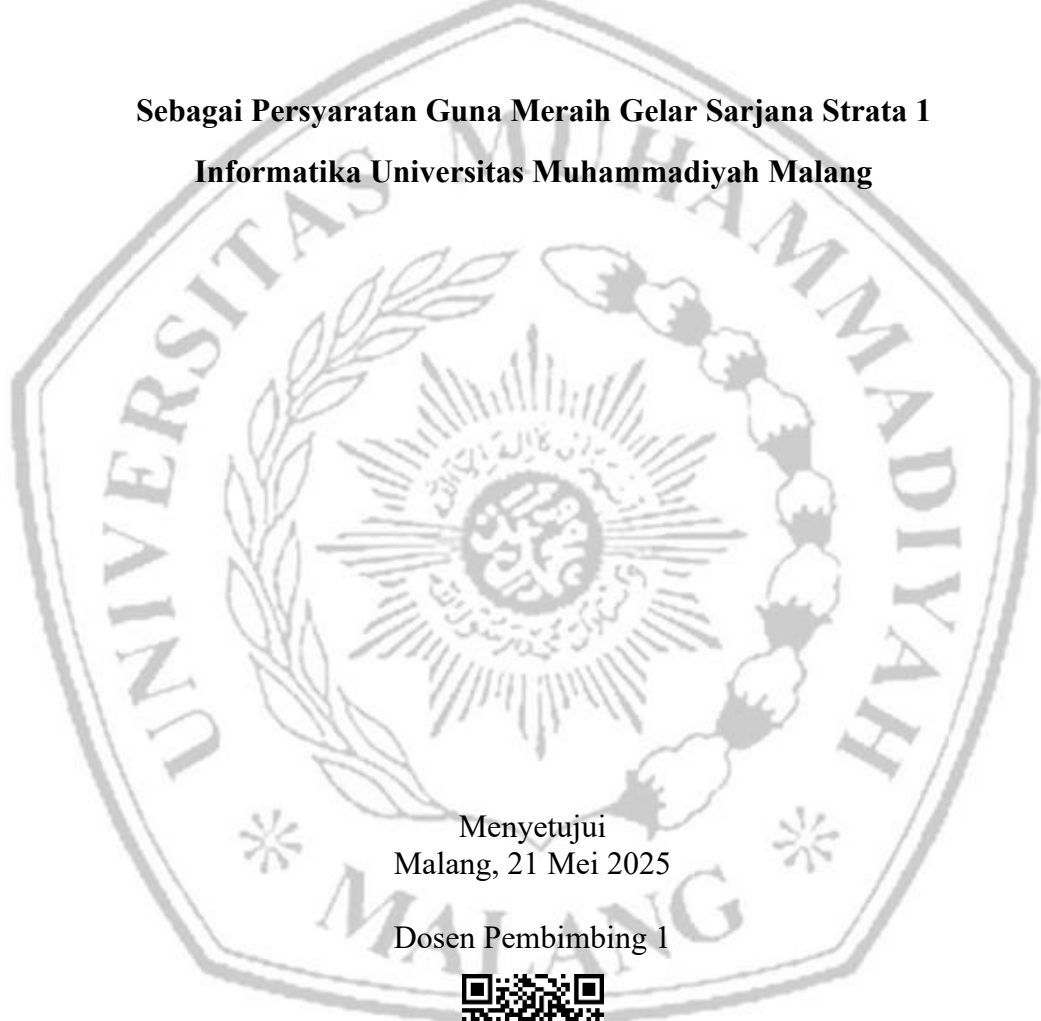
**2025**

## **LEMBAR PERSETUJUAN**

**Deteksi Tren Topik Penelitian Bidang Kecerdasan Buatan Menggunakan  
LDA dan BERTopic**

**TUGAS AKHIR**

**Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang**



Menyetujui  
Malang, 21 Mei 2025

Dosen Pembimbing 1



Setio Basuki M.T., Ph.D.  
NIP. 10809070477PNS

## LEMBAR PENGESAHAN

**Deteksi Tren Topik Penelitian Bidang Kecerdasan Buatan Menggunakan  
LDA & BERTopic**

### TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

**Rahajeng Febri Shafiyah**

**202110370311308**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji  
pada tanggal 26 Juni 2025

Menyetujui,

Dosen Penguji 1



Vinna Rahmayanti S S.si., M.Si  
NIP. 180306071990PNS.

Dosen Penguji 2



Ir. Yufis Azhar S.Kom., M.Kom  
NIP. 10814100544PNS.

Mengetahui,

Ketua Jurusan Informatika



Dr. Galih Wasis Wicaksono S.Kom., M.Kom  
NIP. 10814100541PNS.

## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**NAMA : Rahajeng Febri Shafiyah**

**NIM : 202110370311308**

**FAK./JUR. : Informatika**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “**Deteksi Tren Topik Penelitian Bidang Kecerdasan Buatan Menggunakan LDA dan BERTopic**” beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk risiko/sanksi yang berlaku.

Mengetahui,  
Dosen Pembimbing



Setio Basuki M.T., Ph.D.

Malang, 21 Mei 2025  
Yang Membuat Pernyataan



Rahajeng Febri Shafiyah

## ABSTRAK

Penelitian yang semakin berkembang dari tahun ke tahun membuat para peneliti kesulitan dalam memahami tren topik penelitian tanpa membaca semua publikasi ilmiah. Penelitian ini memiliki tujuan mengetahui tren topik penelitian secara otomatis pada publikasi ICLR (*International Conference On Learning Representation*) dari tahun 2019 hingga 2023 menggunakan LDA (*Latent Dirichlet Allocation*) dan BERTopic. Data dikumpulkan dengan cara crawling, total terdapat 14.613 data. Pengolahan data pada penelitian ini, melalui beberapa tahapan diantaranya preprocessing, pembuatan corpus & dictionary untuk LDA, menerapkan pemodelan LDA dan BERTopic, evaluasi hasil pemodelan menggunakan coherence. Hasil dari penelitian menunjukkan bahwa kedua pemodelan tersebut dapat mengidentifikasi tren topik penelitian setiap tahun, dan pada pemodelan BERTopic memiliki nilai coherence lebih tinggi daripada LDA, sehingga menandakan pemodelan BERTopic memiliki kualitas yang lebih baik dalam menggambarkan setiap kata yang memiliki keterkaitan dalam suatu topik.

**Kata kunci:** Tren topik penelitian, LDA (*Latent Dirichlet Allocation*), BERTopic, coherence

## ABSTRACT

Research that is growing from year to year makes it difficult for research to understand trends in research topics without reading all scientific publications. This study aims to automatically determine trend of research topics in ICLR (International Conference on Learning Representation) from 2019 to 2023 using LDA (Latent Dirichlet Allocation) and BERTopic. Data is collected by crawling, a total of 14.613 data. Data preprocessing in this research, through several stages including preprocessing, corpus & dictionary for LDA, applying LDA and BERTopic modeling, evaluating modelling result using coherence. The result of this research show that both modelling can identify trends in research topics every year, and BERTopic modelling has a higher coherence value than LDA, this indicating that BERTopic has better quality in describing each word that has a relationship in a topic

**Keywords :** Research topic trends, LDA (Latent Dirichlet Allocation), BERTopic, coherence

## LEMBAR PESEMBAHAN

Penulis mengungkapkan puji syukur atas ke hadirat Allah SWT rahmat dan hidayah Nya sehingga penulis berhasil menyelesaikan Tugas Akhir ini dengan baik. Dalam kesempatan ini, penulis ingin mengungkapkan banyak terima kasih kepada semua pihak yang sudah memberikan doa dan dukungan selama proses pengerjaan tugas akhir. Ucapan terima kasih dari penulis ditunjukkan kepada :

1. Diri sendiri yang telah berjuang dari awal hingga tugas akhir ini dapat selesai.
2. Kedua orang tua tercinta, ibu Siti Aisyah dan bapak Ibrahim Faji, atas semua kasih sayang, doa, dan dukungan tiada henti.
3. Bapak Setio Basuki, M.T., Ph.D., selaku dosen pembimbing yang sudah memberikan arahan dan bimbingan selama proses tugas akhir ini.
4. Bapak Ir. Denar Regata Akbi, S.Kom., M.Kom., selaku dosen wali atas semua nasihat akademik dan dukungan sampai saat ini.
5. Seluruh dosen Informatika Universitas Muhammadiyah Malang yang sudah memberikan ilmu dan wawasan selama penulis menjalani perkuliahan.
6. Teman-teman terdekat selama perkuliahan, yaitu Anissa Yulidha Rodiyah, Zahra Sabilla Usman, Muhammad Daffa Nugraha, dan Ahmad Kevin Adhira Zakki yang senantiasa memberikan dukungan disetiap keadaan suka maupun duka
7. Keluarga besar yang terus memberikan semangat dan doa, yaitu R.Aj Nurul Rahmaniyah, Lisa Fitriyah, R.Aj Khatijah Meisyifa, R.Aj Sofiyah, Rhenanti Syahla Safira, R.A Maria Islaha, Maulidina Fajrin, Atria Riza Amelia, Melysa Eka Halimatus Zuhra, R.Aj Rizqia Anisa Putri dan seluruh keluarga besar lainnya yang meskipun tidak disebutkan satu persatu.
8. Teman-teman dari masa SMA yang selalu setia memberikan dukungan hingga saat ini, yaitu Selvy Dwi Safitri, Fairuz Mariana Fadhilah, Nihla Rizkiyah, Qothrun Nada, dan Haza Nabila Putri.

9. Teman-teman satu hunian atau kos yang selalu menemani penulis selama perkuliahan, yaitu Rossa Julfiyatun, Adiyati Nur Afifah, Putri Nursiyam, Nita Yulfarida Arini, Intan Shakila Zurina Villa, dan Nur Rizqiah.



Malang, 21 Mei 2025

*Rahajeng*

Rahajeng Febri Shafiyah

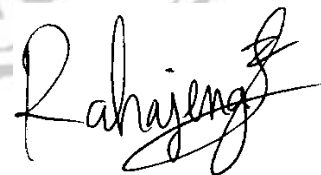
## KATA PENGANTAR

Segala puji bagi Allah SWT atas keberkahan dan hidayah yang diberikan sehingga dapat menyelesaikan penelitian dengan judul “Deteksi Tren Topik Penelitian Bidang Kecerdasan Buatan Menggunakan LDA dan BERTopic”.

Penelitian ini memiliki tujuan untuk mengidentifikasi tren topik penelitian pada publikasi ICLR (*International Conference On Learning Representation*) dari tahun 2019 hingga 2023. Penulis berharap hasil penelitian ini bisa memberikan pengetahuan tentang fase perkembangan penelitian agar para peneliti tidak perlu membaca semua publikasi yang ada.

Pada penulisan ini masih memiliki keterbatasan. Maka dari itu, penulis menerima saran untuk perbaikan dan penyempurnaan supaya penelitian ini memberikan manfaat khususnya kepada para peneliti.

Malang, 21 Mei 2025

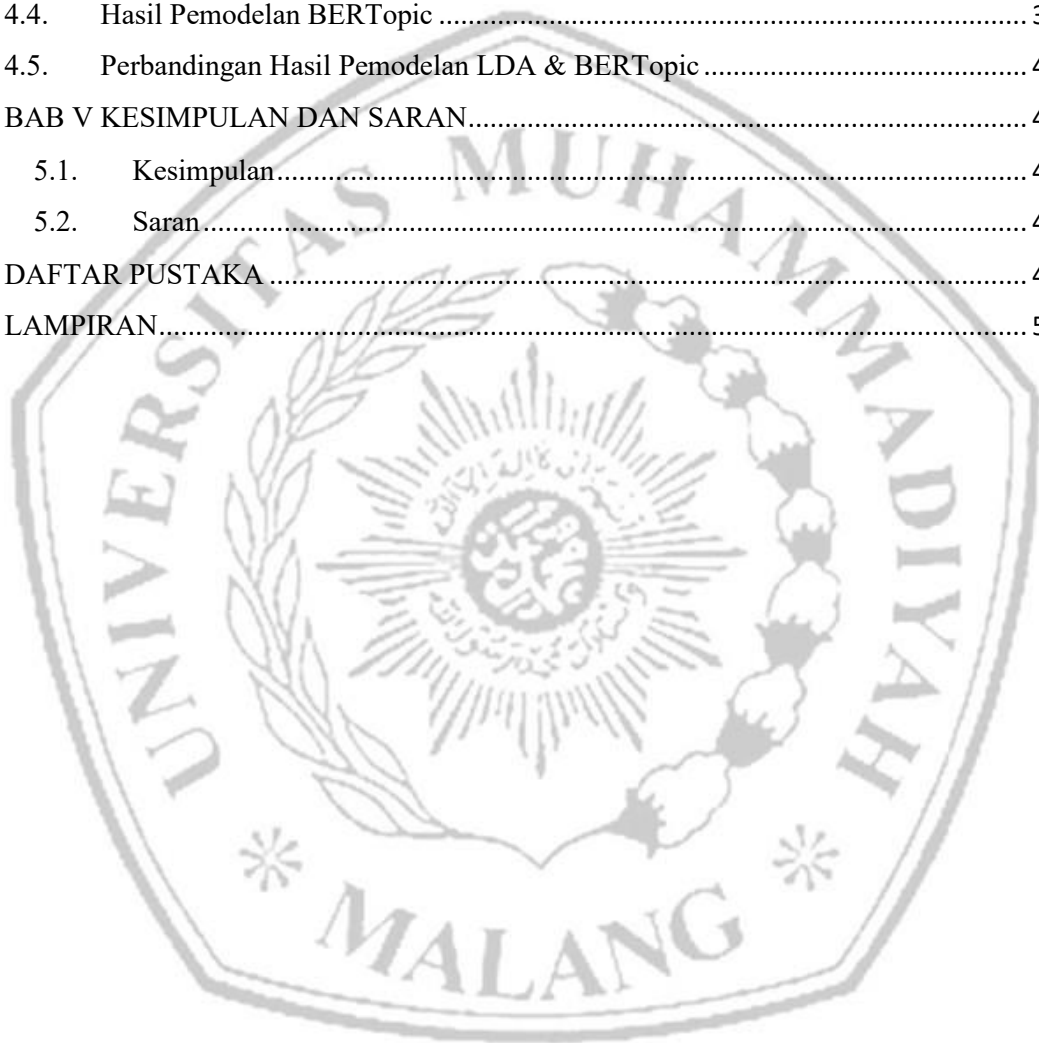


Rahajeng Febri Shafiyah

## DAFTAR ISI

LEMBAR PERSETUJUAN .....	ii
LEMBAR PENGESAHAN .....	iii
LEMBAR PERNYATAAN.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
LEMBAR PESEMBAHAN.....	vii
KATA PENGANTAR .....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah .....	3
1.3. Tujuan Penelitian.....	3
1.4. Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA .....	4
2.1 Studi Literature.....	4
2.2 Pembuatan Corpus & Dictionary.....	5
2.3 Pemodelan LDA .....	6
2.4 Implementasi BERTopic .....	7
2.5 Coherence .....	9
2.6 API Gemini.....	9
BAB III METODOLOGI PENELITIAN .....	10
3.1 Data Penelitian.....	10
3.2 Pengumpulan Data.....	10
3.3 Target Data .....	11
3.4 Tahapan Penelitian .....	12
3.5 Preprocessing.....	13
3.6 Corpus &Dictionary .....	14
3.7 Pembuatan Model LDA.....	14

3.8	Pembuatan Model BERTopic.....	14
3.9	Analisis Fase Perkembangan penelitian .....	15
BAB IV HASIL DAN PEMBAHASAN .....		16
4.1.	Hasil Preprocessing.....	16
4.2.	Pembentukan Corpus & Dictionary .....	16
4.3.	Hasil Pemodelan LDA .....	17
4.4.	Hasil Pemodelan BERTopic .....	36
4.5.	Perbandingan Hasil Pemodelan LDA & BERTopic .....	44
BAB V KESIMPULAN DAN SARAN.....		46
5.1.	Kesimpulan.....	46
5.2.	Saran.....	47
DAFTAR PUSTAKA .....		48
LAMPIRAN.....		51



## DAFTAR GAMBAR

Gambar 2. 1. Pemodelan LDA(sumber Blei et al, 2003).....	6
Gambar 2. 2. Pemodelan BERTopic.....	7
Gambar 3. 1 Distribusi Pada Publikasi ICLR.....	11
Gambar 3. 2 Tahapan penelitian algoritma LDA dan BERTopic.....	12
Gambar 4. 1 Visualisasi Fase Perkembangan Penelitian Menggunakan LDA.....	25
Gambar 4. 2. Visualisasi Word Cloud ICLR 2019 LDA.....	26
Gambar 4. 3. Visualisasi Word Cloud ICLR 2020 LDA.....	26
Gambar 4. 4. Visualisasi Word Cloud ICLR 2021 LDA.....	27
Gambar 4. 5. Visualisasi Word Cloud ICLR 2022 LDA.....	28
Gambar 4. 6. Visualisasi Word Cloud ICLR 2023 LDA.....	28
Gambar 4. 7. Visualisasi Fase Perkembangan Penelitian Menggunakan BERTopic.....	38
Gambar 4. 8. Visualisasi clustering ICLR 2019.....	39
Gambar 4. 9. Visualisasi clustering ICLR 2020.....	40
Gambar 4. 10. Visualisasi clustering ICLR 2021.....	41
Gambar 4. 11. Visualisasi clustering ICLR 2022.....	42
Gambar 4. 12. Visualisasi clustering ICLR 2023.....	43
Gambar 4. 13. Visualiasi perbandingan nilai coherence.....	44

## DAFTAR TABEL

Tabel 3. 1 Sample Dataset ICLR 2023 .....	10
Tabel 4. 1 Hasil Preprocessing Data .....	16
Tabel 4. 2 Hasil Pembentukan Corpus & Dictionary.....	17
Tabel 4. 3 Hasil Pemodelan LDA .....	19
Tabel 4. 4. Validasi Topik Utama ICLR 2019 .....	29
Tabel 4. 5. Validasi Topik Utama ICLR 2020 .....	31
Tabel 4. 6. Validasi Topik Utama ICLR 2021 .....	32
Tabel 4. 7. Validasi Topik Utama ICLR 2022 .....	34
Tabel 4. 8. Validasi Topik Utama ICLR 2023 .....	35
Tabel 4. 9. Hasil Pemodelan BERTopic .....	36
Tabel 4. 10. Perbandingan Nilai Coherence .....	44
Tabel 4. 11. Perbandingan Topik Utama .....	45



## DAFTAR LAMPIRAN

Lampiran 1. Output topic modelling ICLR 2023..... 51

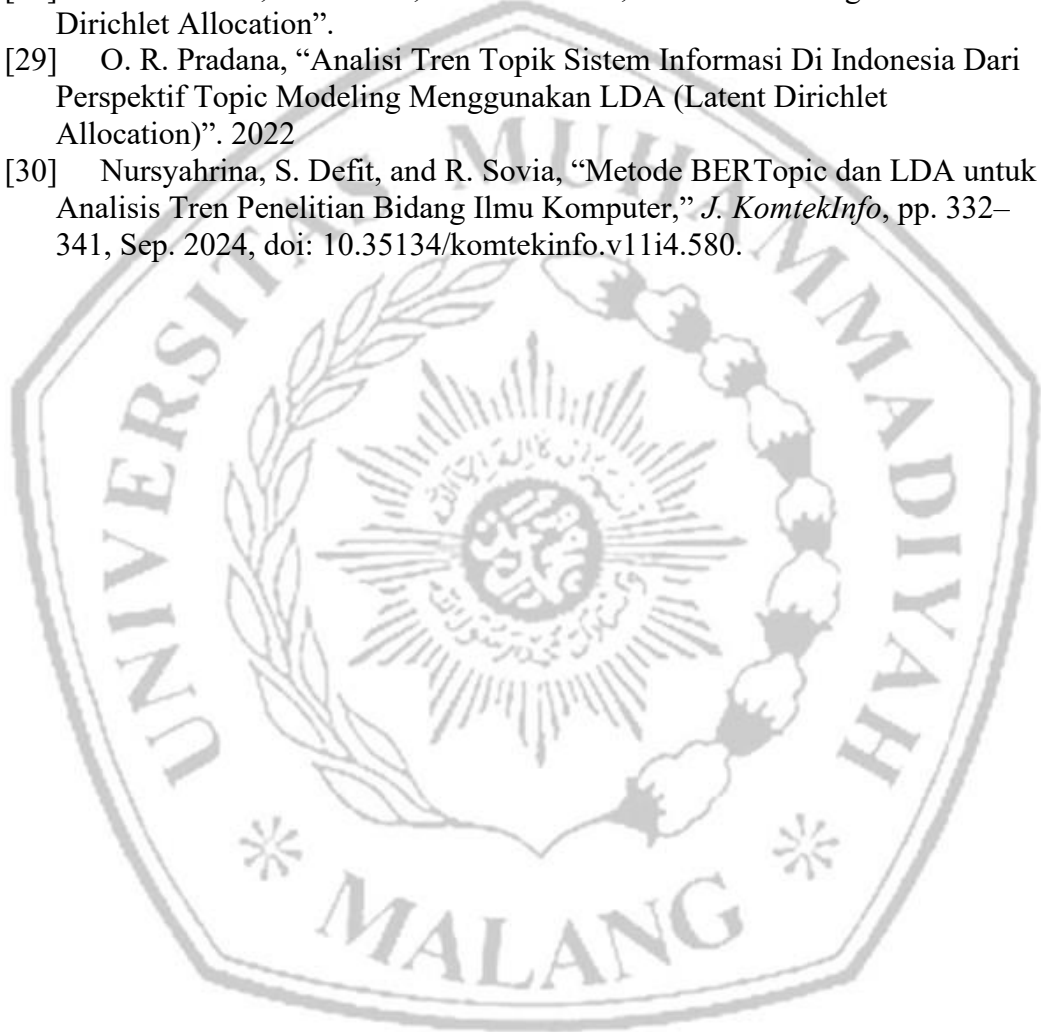


## DAFTAR PUSTAKA

- [1] F. Jeka, R. Risnita, M. S. Jailani, and A. Asrulla, "Kajian Literatur dalam Menyusun Referensi Kunci, State Of The Art, dan Keterbaharuan Penelitian (Novelty)," *J. Pendidik. Tambusai*, vol. 7, Nov. 2023, doi: <https://doi.org/10.31004/jptam.v7i3.10870>.
- [2] C. Li, Y. Chen, and Y. Shang, "A review of industrial big data for decision making in intelligent manufacturing," *Eng. Sci. Technol. Int. J.*, vol. 29, p. 101021, May 2022, doi: 10.1016/j.jestech.2021.06.001.
- [3] H. Jelodar *et al.*, "Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey," Dec. 05, 2018, *arXiv*: arXiv:1711.04305. Accessed: Sep. 25, 2024. [Online]. Available: <http://arxiv.org/abs/1711.04305>
- [4] H. Axelborn and J. Berggren, "Topic Modeling for Customer Insights A Comparative Analysis of LDA and BERTopic in Categorizing Customer Calls," Umeå University, 2023. [Online]. Available: <https://www.bing.com/ck/a?!&&p=8320d47a58d34d242f3aef163a68428adc261b9250d9988e64f952a8216b4826JmltdHM9MTczNjU1MzYwMA&ptn=3&ver=2&hsh=4&fclid=120b6770-653b-6eae-1be8-77c8645e6f65&psq=Topic+Modeling+for+Customer+Insights+A+Comparative+Analysis+of+LDA+and+BERTopic+in+Categorizing+Customer+Calls&u=a1aHR0cHM6Ly91bXUuZGl2YS1wb3J0YWwub3JnL3NtYXNoL2dldC9kaXZhMjoxNzYzNjM3L0ZVTEhUMDEucGRm&ntb=1>
- [5] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," Mar. 11, 2022, *arXiv*: arXiv:2203.05794. Accessed: Aug. 19, 2024. [Online]. Available: <http://arxiv.org/abs/2203.05794>
- [6] Y. Barkette and P. Bailette, "Big data analytics in turbulent contexts: towards organizational change for enhanced agility," *Prod. Plan. Control*, vol. 33(2–3), pp. 105–122, 2020.
- [7] Y. Liu and F. Wan, "Unveiling temporal and spatial research trends in precision agriculture: A BERTopic text mining approach," *Heliyon*, vol. 10, no. 17, p. e36808, Sep. 2024, doi: 10.1016/j.heliyon.2024.e36808.
- [8] D. Atzeni, D. Bacciu, D. Mazzei, and G. Prencipe, "A Systematic Review of Wi-Fi and Machine Learning Integration with Topic Modeling Techniques," *Sensors*, vol. 22, no. 13, p. 4925, Jun. 2022, doi: 10.3390/s22134925.
- [9] A. Erna and A. K. Jailani, "Identifikasi Potensi Topik Penelitian Baru: Analisis Judul Skripsi Teknik Informatika (2022-2024) menggunakan Latent Dirichlet Allocation (LDA)," no. 2, 2024.
- [10] M. Adewunmi, S. K. Sharma, N. Sharma, N. S. Sharma, and B. Mounmo, "Cancer Health Disparities drivers with BERTopic modelling and PyCaret Evaluation".
- [11] Y. Sahria and D. H. Fudholi, "Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (Latent Dirichlet Allocation)," *Vol .*, no. 2, 2020.
- [12] 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," *Glob.*

- Transit. Proc.*, vol. 3, no. 1, pp. 298–304, Jun. 2022, doi: 10.1016/j.gltp.2022.03.015.
- [13] S. H. Mohammed and S. Al-augby, “LSA & LDA topic modeling classification: comparison study on e-books,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 1, p. 353, Jul. 2020, doi: 10.11591/ijeecs.v19.i1.pp353-362.
- [14] D. M. Blei, “Latent Dirichlet Allocation”.
- [15] A. F. Pathan and C. Prakash, “Unsupervised Aspect Extraction Algorithm for opinion mining using topic modeling,” *Glob. Transit. Proc.*, vol. 2, no. 2, pp. 492–499, Nov. 2021, doi: 10.1016/j.gltp.2021.08.005.
- [16] K. B. Vamshi, A. K. Pandey, and K. A. P. Siva, “Topic Model Based Opinion Mining and Sentiment Analysis,” in *2018 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore: IEEE, Jan. 2018, pp. 1–4. doi: 10.1109/ICCCI.2018.8441220.
- [17] D. Maulidiya, “Topic Modelling using Latent Dirichlet Allocation (LDA) to Investigate the Latent Topics of Mathematical Creative Thinking Research in Indonesia,” *J. Intell. Comput. Health Inform.*, vol. 3, no. 2, p. 35, Feb. 2023, doi: 10.26714/jichi.v3i2.11428.
- [18] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, “Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 1, pp. 123–131, Feb. 2021, doi: 10.29207/resti.v5i1.2587.
- [19] A. R. Hadiat, “Topic Modeling Evaluations: The Relationship Between Coherency and Accuracy”.
- [20] R. Egger and J. Yu, “A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts,” *Front. Sociol.*, vol. 7, p. 886498, May 2022, doi: 10.3389/fsoc.2022.886498.
- [21] S. Samsir, R. S. Saragih, S. Subagio, R. Aditiya, and R. Watrianthos, “BERTopic Modeling of Natural Language Processing Abstracts: Thematic Structure and Trajectory,” *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 3, p. 1514, Jul. 2023, doi: 10.30865/mib.v7i3.6426.
- [22] L. McInnes, J. Healy, and J. Melville, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction,” 2018, *arXiv*. doi: 10.48550/ARXIV.1802.03426.
- [23] V. Sharifian-Attar, S. De, S. Jabbari, J. Li, H. Moss, and J. Johnson, “Analysing Longitudinal Social Science Questionnaires: Topic modelling with BERT-based Embeddings,” in *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan: IEEE, Dec. 2022, pp. 5558–5567. doi: 10.1109/BigData55660.2022.10020678.
- [24] S. Aprilia, R. Agustin, V. H. Pranatawijaya, and N. N. K. Sari, “PENERAPAN API GEMINI DALAM LAYANAN PEMINJAMAN NOVEL ONLINE PADA WEBSITE COZYBOOK,” *J. Inform. Dan Tek. Elektro Terap.*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4508.
- [25] S. Zhou, P. Kan, Q. Huang, and J. Silbernagel, “A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during

- Hurricane Laura,” *J. Inf. Sci.*, vol. 49, no. 2, pp. 465–479, Apr. 2023, doi: 10.1177/01655515211007724.
- [26] A. Ariansyah and U. Indahyanti, “Fitur Ekstraksi pada Pemodelan Topik Menggunakan Metode Latent Dirichlet Allocation pada Peristiwa Kebocoran Data,” vol. 1, no. 2, 2024.
- [27] M. Sharipov and O. Sobirov, “Development of a rule-based lemmatization algorithm through Finite State Machine for Uzbek language,” 2022, *arXiv*. doi: 10.48550/ARXIV.2210.16006.
- [28] M. Hoffman, F. R. Bach, and D. M. Blei, “Online Learning for Latent Dirichlet Allocation”.
- [29] O. R. Pradana, “Analisi Tren Topik Sistem Informasi Di Indonesia Dari Perspektif Topic Modeling Menggunakan LDA (Latent Dirichlet Allocation)”. 2022
- [30] Nursyahrina, S. Defit, and R. Sovia, “Metode BERTopic dan LDA untuk Analisis Tren Penelitian Bidang Ilmu Komputer,” *J. KomtekInfo*, pp. 332–341, Sep. 2024, doi: 10.35134/komtekinfo.v1i1i4.580.



### FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

**Nama Mahasiswa** : Rahajeng Febri Shafiyah  
**NIM** : 202110370311308  
**Judul TA** : Deteksi Tren Topik Penelitian Bidang Kecerdasan Buatan Menggunakan LDA dan BERTopic

#### Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	7 %
2.	Bab 2 – Daftar Pustaka	25 %	2 %
3.	Bab 3 – Analisis dan Perancangan	25 %	2 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	2 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	3 %
6.	Makalah Tugas Akhir	20%	0 %

\*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

\*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)



(.....*Rahajeng Febri Shafiyah*.....)