

Implementasi Algoritma Topic Modeling pada Abstrak Paper Ilmiah untuk Deteksi Tren Topik Penelitian

Laporan Tugas Akhir

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



Nurul Maudy Fadillah
202010370311445

Bidang Minat :
Data Science

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2024**

LEMBAR PERSETUJUAN

**Implementasi Algoritma Topic Modeling pada Abstrak Paper Ilmiah untuk
Deteksi Tren Topik Penelitian**

TUGAS AKHIR

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

Menyetujui,

Malang, 19 Juli 2024

Dosen Pembimbing 1



Setio Basuki MT., Ph.D.

NIP. 10809070477PNS.

LEMBAR PENGESAHAN

**Implementasi Algoritma Topic Modeling pada Abstrak Paper
Ilmiah untuk Deteksi Tren Topik Penelitian**

TUGAS AKHIR

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

Disusun Oleh :

Nurul Maudy Fadillah

202010370311445

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji
pada tanggal 19 Juli 2024

Menyetujui,

Dosen Penguji 1



Ir. Yufis Azhar S.Kom., M.Kom.

NIP. 10814100544PNS.

Dosen Penguji 2



Yuda Munarko S.Kom., M.Sc

NIP. 10806110443PNS.

Mengetahui,

Ketua Jurusan Informatika



Ir. Gafih Wasis Wicaksono S.kom. M.Cs.
NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : Nurul Maudy Fadillah

NIM : 202010370311445

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“Implementasi Algoritma Topic Modeling pada Abstrak Paper Ilmiah untuk Deteksi Tren Topik Penelitian”** 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 resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing



Setio Basuki MT., Ph.D.

Malang, 19 Juli 2024
Yang Membuat Pernyataan



Nurul Maudy Fadillah

ABSTRAK

Memahami perkembangan dan fokus penelitian dalam bidang bahasa serta evaluasi sumber daya bahasa sangat penting untuk mendukung penelitian lebih lanjut. Penelitian ini bertujuan untuk mendeteksi tren topik penelitian menggunakan algoritma Topic Modeling pada abstrak paper ilmiah. Metode yang diterapkan adalah Latent Dirichlet Allocation (LDA), yang memungkinkan identifikasi topik secara otomatis dari kumpulan dokumen. Proses penelitian melibatkan beberapa tahap, termasuk pengumpulan data, preprocessing, pembobotan topik, dan analisis hasil model LDA. Model LDA kemudian digunakan untuk mengelompokkan data menjadi beberapa topik yang dominan. Hasil penelitian menunjukkan bahwa metode LDA efektif dalam mengidentifikasi tren topik penelitian. Pada model LDA dengan 3 topik, mencerminkan seberapa rumitnya proses pengolahan bahasa alami, dari sisi teknis hingga upaya pengembangan sumber daya dan penelitian linguistik. Sedangkan pada model LDA dengan 5 topik, memberikan gambaran tentang analisis teks dan bahasa. Penelitian ini diharapkan dapat memberikan kontribusi dalam bidang analisis teks dan pemodelan topik, serta mendukung peneliti dalam mengidentifikasi tren terbaru dalam literatur ilmiah.

Kata Kunci: Topic Modeling, Latent Dirichlet Allocation (LDA), Abstrak Paper Ilmiah, Tren Penelitian, Analisis Teks.

ABSTRACT

Understanding the development and research focus in the field of language, as well as evaluating language resources, are crucial to support further research endeavors. This study aims to detect research topic trends using the Topic Modeling algorithm on abstracts of scientific papers. The method applied is Latent Dirichlet Allocation (LDA), enabling automatic topic identification from a document corpus. The research process involves several stages, including data collection, preprocessing, topic weighting, and analysis of LDA model outcomes. The LDA model is then utilized to cluster data into dominant topics. The findings indicate the effectiveness of the LDA method in identifying research topic trends. In the LDA model with 3 topics, it reflects the complexity of natural language processing, from technical aspects to efforts in resource development and linguistic research. Meanwhile, in the LDA model with 5 topics, it provides an overview of text and language analysis. This research is anticipated to contribute to the fields of text analysis and topic modeling, aiding researchers in identifying recent trends in scientific literature.

Keywords: Topic Modeling, Latent Dirichlet Allocation (LDA), Scientific Paper Abstract, Research Trends, Text Analysis.

LEMBAR PERSEMBAHAN

Puji dan syukur penulis haturkan kepada Allah SWT atas limpahan rahmat dan karunia-Nya yang telah memberikan kemampuan kepada penulis untuk menyelesaikan Tugas Akhir ini. Penulis juga ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT sebagai pencipta seluruh alam, telah memberikan kesehatan dan kelancaran kepada penulis dalam mengerjakan penelitian ini.
2. Kedua orang tua penulis Bapak Mansyur dan Ibu Sri Mulyani yang senantiasa memberikan dorongan, semangat, kasih sayang, doa, dan dukungan sepenuhnya kepada penulis dalam mengerjakan penelitian ini.
3. Kakak dan adik penulis Ratih Herliana dan Muhammad Ade Rizki yang selalu memberikan semangat, dukungan, kasih sayang, dan doa kepada penulis dalam mengerjakan penelitian ini.
4. Dosen pembimbing penulis Bapak Setio Basuki MT., Ph.D yang telah meluangkan waktu untuk membimbing penulis sehingga dapat menyelesaikan penelitian tugas akhir ini.
5. Dosen wali penulis Bapak Luqman Hakim, S.Kom., M.Kom yang telah memberikan bimbingan sejak awal perkuliahan hingga saat ini.
6. Dosen dan staf di Program Studi Informatika Universitas Muhammadiyah Malang telah memberikan pengetahuan selama masa kuliah dan memberikan bantuan yang sangat berarti kepada penulis.
7. Saya pribadi yang telah berhasil menyelesaikan studi.
8. Teman terdekat penulis Abrar Riner Lawdza'i, Siti Nurlaili Oktafiani, Putri Nurul Umni, Zahrawana Putri Yodita, dan Hania Pratiwi Ningrum yang selalu memberi semangat dan motivasi untuk menyelesaikan penelitian ini.

Malang, 12 Juni 2024



Nurul Maudy Fadillah

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Allah SWT. Atas limpahan rahmat dan hidayah-NYA sehingga peneliti dapat menyelesaikan penelitian dengan judul "Implementasi Algoritma Topic Modeling pada Abstrak Paper Ilmiah untuk Deteksi Tren Topik Penelitian".

Penelitian ini dilakukan dengan tujuan untuk menggali dan menganalisis tren topik penelitian dalam abstrak paper ilmiah menggunakan algoritma Topic Modeling. Proses ini diharapkan dapat memberikan wawasan mendalam tentang perkembangan dan fokus penelitian dalam suatu domain tertentu. Sebagai penulis menyadari bahwa penulisan ini tidak terlepas dari berbagai keterbatasan, namun dengan kerendahan hati, penulis menyampaikan hasil penelitian ini dengan harapan dapat memberikan kontribusi positif terutama dalam memahami dinamika topik penelitian saat ini.

Peneliti menyadari sepenuhnya bahwa tugas akhir ini memiliki kekurangan dan keterbatasan. Oleh karena itu, peneliti mengharapkan masukan agar tulisan ini dapat memberikan manfaat bagi kemajuan ilmu pengetahuan.

Malang, 12 Juni 2024



Nurul Maudy Fadillah

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1. Pendahuluan.....	1
1.2. Rumusan Masalah.....	2
1.3. Tujuan Penelitian.....	3
1.4. Batasan Masalah.....	3
BAB II STUDI LITERATUR	4
BAB III METODOLOGI PENELITIAN	8
3.1. Populasi dan Sampel Penelitian.....	8
3.2. Variabel dan Definisi Operasional Penelitian	8
3.3. Rencana penelitian.....	8
3.4. Pengumpulan Data.....	9
3.5. Preprocessing data	9
3.6. Pembobotan Topik Dominan.....	10
3.7. Pembuatan Model LDA.....	10

3.8. Pembentukan Topik.....	10
3.9. Analisis Tren.....	11
BAB IV HASIL DAN PEMBAHASAN	12
4.1. Hasil Pengumpulan Data	12
4.2. Hasil Preprocessing Data.....	12
4.3. Pembobotan Topik Dominan.....	13
4.4. Hasil model LDA.....	15
4.4.1. Model LDA Cluster 3 Topic	15
4.4.2. Model LDA Cluster 5 Topic	20
BAB V KESIMPULAN DAN SARAN.....	26
5.1. Kesimpulan.....	26
5.2. Saran.....	26
DAFTAR PUSTAKA.....	28
LAMPIRAN.....	31



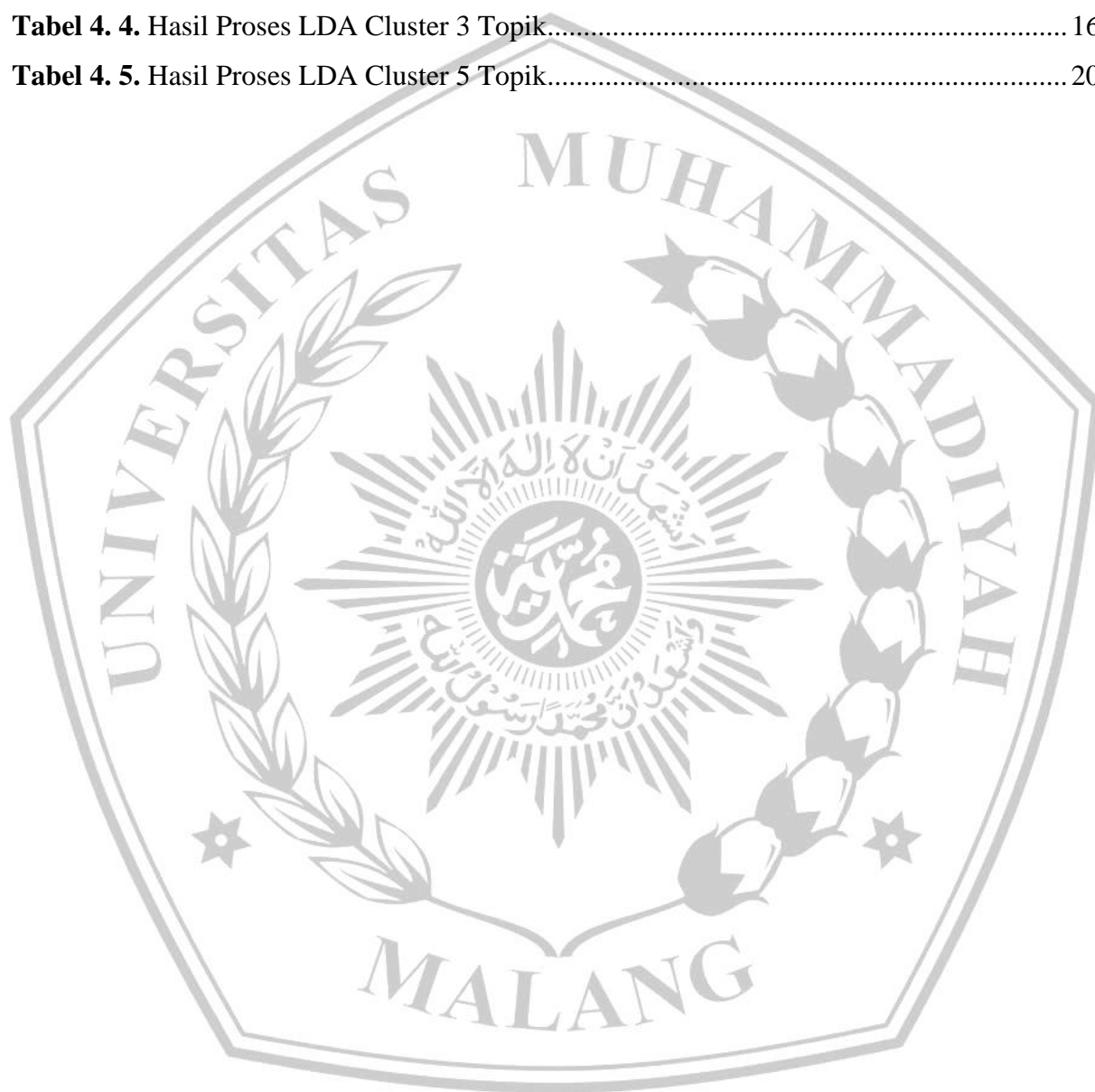
DAFTAR GAMBAR

Gambar 3. 1. Tahap Penelitian	8
Gambar 4. 1. Jumlah Abstrak Paper Ilmiah	12
Gambar 4. 2. Word Clouds dan Word Counts Cluster 3 Topik	17
Gambar 4. 3. PyLDAvis Cluster 3 Topik.....	19
Gambar 4. 4. Word Clouds dan Word Counts Cluster 5 Topik	22
Gambar 4. 5. PyLDAvis Cluster 5 Topik.....	24



DAFTAR TABEL

Tabel 3. 1. Definisi Operasional Penelitian.....	8
Tabel 4. 1. Hasil Preprocessing Data	12
Tabel 4. 2. Hasil Pembobotan Cluster 3 Topik	14
Tabel 4. 3. Hasil Pembobotan Cluster 5 Topik	14
Tabel 4. 4. Hasil Proses LDA Cluster 3 Topik.....	16
Tabel 4. 5. Hasil Proses LDA Cluster 5 Topik.....	20



DAFTAR LAMPIRAN

Lampiran 1. Output Topic Modeling	31
Lampiran 2. Sertifikat Plagiasi.....	32



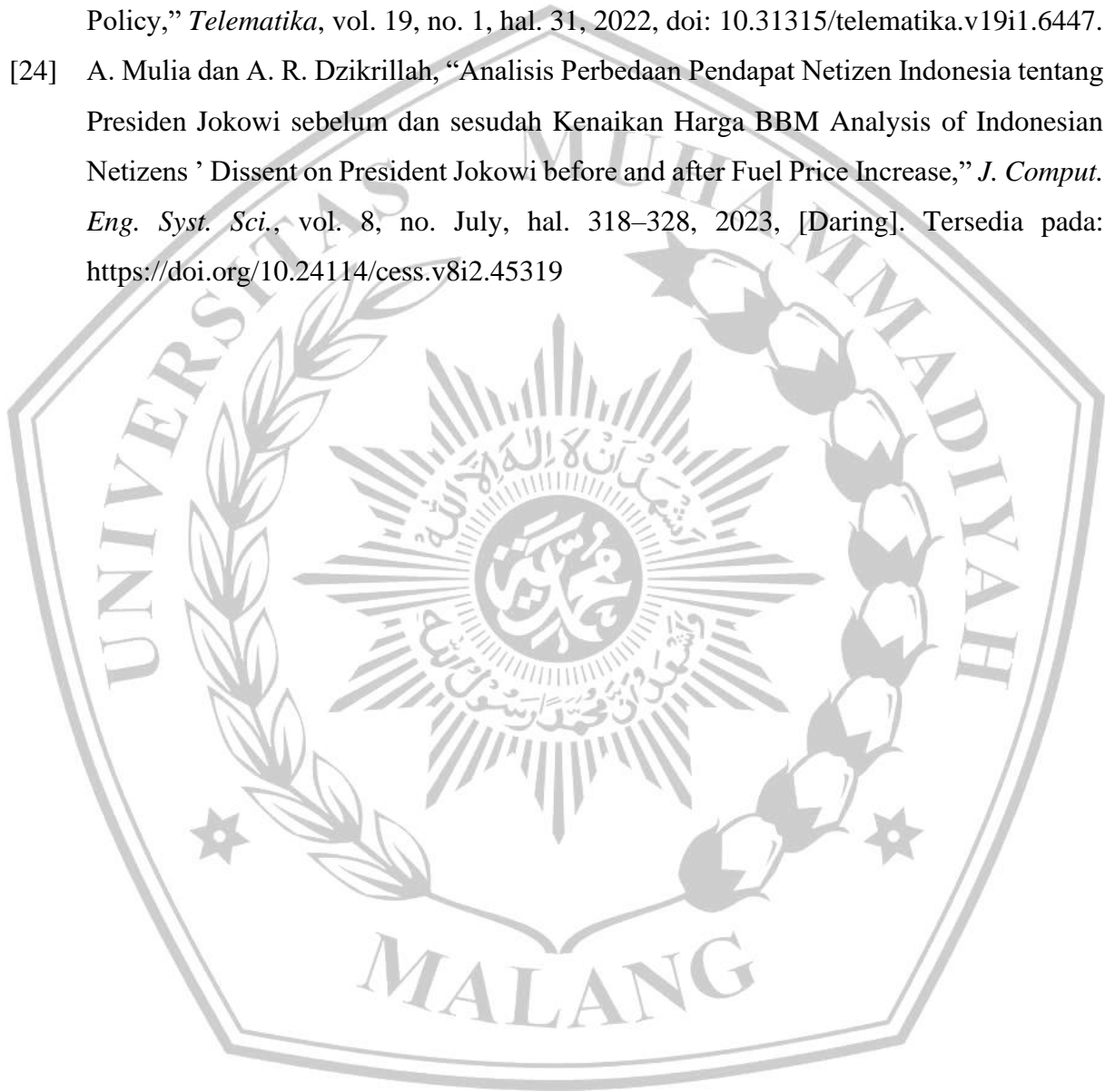
DAFTAR PUSTAKA

- [1] M. K. M. Nasution, “Abstrak: Suatu karya ilmiah,” *Researchgate*, no. December, hal. 1–4, 2017, doi: 10.13140/RG.2.2.14333.26086.r
- [2] E. L. NAPITUPULU, “Kualitas Publikasi Ilmiah Internasional Indonesia Terus Didongkrak,” *kompas.id*, 2022, [Daring]. Tersedia pada: <https://www.kompas.id/baca/humaniora/2022/11/17/publikasi-ilmiah-internasional-indonesia-terus-didongkrak>
- [3] A. I. Alfanzar dan I. S. Rozas, “Topic Modelling Skripsi Menggunakan Metode Latent,” *JSiI (Jurnal Sist. Informasi)*, vol. 7, no. 1, hal. 7–13, 2020.
- [4] N. A. S. ER, “Implementasi Latent Dirichlet Allocation (LDA) untuk Klasterisasi Cerita Berbahasa Bali,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, hal. 127–134, 2021, doi: 10.25126/jtiik.202183556.
- [5] M. R. Fahlevvi dan A. SN, “Topic Modeling on Online News.Portal Using Latent Dirichlet Allocation (LDA),” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 16, no. 4, hal. 335, 2022, doi: 10.22146/ijccs.74383.
- [6] J. W. Fernanda, I. Agama, I. Negeri, dan I. Kediri, “PEMODELAN PERSEPSI PEMBELAJARAN ONLINE MENGGUNAKAN LATENT DIRICHLET ALLOCATION,” vol. 9, no. 2, 2021.
- [7] T. Informatika dan U. A. Yogyakarta, “Pengamatan tren ulasan hotel menggunakan pemodelan topik berbasis latent dirichlet allocation,” vol. 6, no. 2, hal. 169–175, 2022.
- [8] R. Sistem, “JURNAL RESTI Sentiment Analysis of Public Opinion Related to Rapid Test Using LDA,” no. 10, hal. 672–679, 2021.
- [9] I. Z. Rakhmat Kurniawan R, “Penerapan Text Mining Pada Sistem Penyeleksian Judul Skripsi Menggunakan Algoritma Latent Dirichlet Allocation (LDA),” vol. 11, hal. 1036, 2022, [Daring]. Tersedia pada: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3120/84>
- [10] Y. Sahria dan D. Hatta Fudholi, “Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (Latent Dirichlet Allocation),” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 3, hal. 336–344, 2017.
- [11] I. N. Kabiru dan P. K. Sari, “Analisa Konten Media Sosial E-commerce Pada Instagram Menggunakan Metode Sentiment Analysis Dan Lda-based Topic Modeling (studi Kasus: Shopee Indonesia),” *eProceedings Manag.*, vol. 6, no. 1, hal. 12–19, 2019, [Daring]. Tersedia pada:

<https://openlibrarypublications.telkomuniversity.ac.id/index.php/management/article/view/8498>

- [12] B. W. Arianto dan G. Anuraga, "Topic Modeling for Twitter Users Regarding the 'Ruangguru' Application," *J. ILMU DASAR*, vol. 21, no. 2, hal. 149, 2020, doi: 10.19184/jid.v21i2.17112.
- [13] N. Cahyono dan Angga Reni Dwi Astuti, "Analisis Topic Modelling Persepsi Pengguna Internet Menggunakan Metode Latent Dirichlet Allocation," *Indones. J. Comput. Sci.*, vol. 12, no. 1, hal. 326–334, 2023, doi: 10.33022/ijcs.v12i1.3155.
- [14] H. Sagala dan H. Toba, "Penentuan Aspek yang Berpengaruh Terhadap Produk Smartphone Berdasarkan Ulasan Berbasis Teksual," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, hal. 287–295, 2021, doi: 10.28932/jutisi.v7i1.3466.
- [15] Kamdan, Ivana Lucia Kharisma, Gina Purnama Insany, dan Paikun, "Research topic modeling in informatics engineering study program at Nusa Putra University using LDA method," *Int. J. Eng. Appl. Technol.*, vol. 5, no. 2, hal. 24–35, 2022, doi: 10.52005/ijeat.v5i2.76.
- [16] E. Gryaznova dan M. Kirina, "Defining Kinds of Violence in Russian Short Stories of 1900-1930: A Case of Topic Modelling With LDA and PCA," *CEUR Workshop Proc.*, vol. 3090, hal. 281–290, 2021.
- [17] W. A. Nawang Sari dan H. Dwi Purnomo, "Topic Modeling Using the Latent Dirichlet Allocation Method on Wikipedia Pandemic Covid-19 Data in Indonesia," *J. Tek. Inform.*, vol. 3, no. 5, hal. 1223–1230, 2022, doi: 10.20884/1.jutif.2022.3.5.321.
- [18] H. Fitri, W. Widyawan, dan I. Soesanti, "Topic Modeling in the News Document on Sustainable Development Goals," *IJITEE (International J. Inf. Technol. Electr. Eng.)*, vol. 5, no. 3, hal. 82, 2021, doi: 10.22146/ijitee.67467.
- [19] O. Ozyurt, "Empirical research of emerging trends and patterns across the flipped classroom studies using topic modeling," *Educ. Inf. Technol.*, vol. 28, no. 4, hal. 4335–4362, 2023, doi: 10.1007/s10639-022-11396-8.
- [20] I. N. Switrayana, D. Ashadi, H. Hairani, dan A. Aminuddin, "Sentiment Analysis and Topic Modeling of Kitabisa Applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods," *Int. J. Eng. Comput. Sci. Appl.*, vol. 2, no. 2, hal. 81–91, 2023, doi: 10.30812/ijecsa.v2i2.3406.
- [21] Janu Akrama Wardhana dan Yuliant Sibaroni, "Aspect Level Sentiment Analysis on Zoom Cloud Meetings App Review Using LDA," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, hal. 631–638, 2021, doi: 10.29207/resti.v5i4.3143.

- [22] M. R. Irianto, A. Maududie, dan F. N. Arifin, "Implementation of K-Means Clustering Method for Trend Analysis of Thesis Topics (Case Study: Faculty of Computer Science, University of Jember)," *Berk. Sainstek*, vol. 10, no. 4, hal. 210, 2022, doi: 10.19184/bst.v10i4.29524.
- [23] M. Habibi, M. R. Ma'arif, dan D. Subekti, "The Development of Social Media Intelligence System for Citizen Opinion and Perception Analysis over Government Policy," *Telematika*, vol. 19, no. 1, hal. 31, 2022, doi: 10.31315/telematika.v19i1.6447.
- [24] A. Mulia dan A. R. Dzikrillah, "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," *J. Comput. Eng. Syst. Sci.*, vol. 8, no. July, hal. 318–328, 2023, [Daring]. Tersedia pada: <https://doi.org/10.24114/cess.v8i2.45319>





UNIVERSITAS
MUHAMMADIYAH
MALANG



FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Nurul Maudy Fadillah
NIM : 202010370311445
Judul TA : Implementasi Algoritma Topic Modeling pada Abstrak Paper Ilmiah untuk Deteksi Tren Topik Penelitian

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	0 %
2.	Bab 2 – Studi Literatur	25 %	5 %
3.	Bab 3 – Metodologi Penelitian	25 %	2 %
4.	Bab 4 – Hasil dan Pembahasan	15 %	3 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	2 %
6.	Makalah Tugas Akhir	20 %	0 %

*) Hasil cek plagiarisme diisi oleh pemeriksa (staf TU)

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

Mengetahui,

Pemeriksa (Staff TU)

(.....)



Kampus I

Jl. Bandung 1 Malang, Jawa Timur
P. +62 341 551 253 (Hunting)
F. +62 341 460 435

Kampus II

Jl. Bendungan Sutarni No 186 Malang, Jawa Timur
P. +62 341 551 149 (Hunting)
F. +62 341 582 080

Kampus III

Jl. Raya Tlogomas No 246 Malang, Jawa Timur
P. +62 341 464 318 (Hunting)
F. +62 341 460 435
E: webmaster@umm.ac.id