

**Implementasi Algoritma Topic Modeling pada Data Cuitan
Terkait Kesehatan Mental pada Platform Media Sosial Twitter
(X) Berbahasa Indonesia**

Laporan Tugas Akhir

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



Nico Ardia Effendy
202010370311467

Bidang Minat
Data Science

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

2024

LEMBAR PERSETUJUAN

**Implementasi Algoritma Topic Modeling pada Data Cuitan
Terkait Kesehatan Mental pada Platform Media Sosial Twitter
(X) Berbahasa Indonesia**

TUGAS AKHIR

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



Menyetujui,

Malang, 24 September 2024

Dosen Pembimbing 1



Setio Basuki MT., Ph.D.

NIP. 10809070477PNS.

LEMBAR PENGESAHAN

**Implementasi Algoritma Topic Modeling pada Data Cuitan
Terkait Kesehatan Mental pada Platform Media Sosial Twitter
(X) Berbahasa Indonesia**

TUGAS AKHIR

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

Disusun Oleh :

Nico Ardia Effendy

202010370311467

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji
pada tanggal 24 September 2024

Menyetujui,

Dosen Penguji 1



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

NIP. 10814100544PNS.

Dosen Penguji 2



Vinna Rahmayanti S S.Si., M.Si

NIP. 180306071990PNS.

Mengetahui,
Ketua Jurusan Informatika



Ir. Galih Wasis Wicaksono S.kom. M.Cs.

NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : Nico Ardia Effendy

NIM : 202010370311467

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“Implementasi Algoritma Topic Modeling pada Data Cuitan Terkait Kesehatan Mental pada Platform Media Sosial Twitter (X) Berbahasa Indonesia”** 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, 24 September 2024
Yang Membuat Pernyataan



Nico Ardia Effendy

ABSTRAK

Penelitian ini bertujuan untuk mengimplementasikan algoritma *topic modeling* pada data cuitan teks kesehatan mental, dengan meninjau literatur mengenai topik ini mengungkapkan bahwa penelitian sebelumnya memiliki fokus yang sempit, menyebabkan banyak isu terabaikan. Makalah ini membahas masalah tersebut dengan menerapkan metode *topic modeling* di berbagai kerangka kesehatan mental melalui beberapa tahap. Pertama, tulisan ini mengkaji berbagai bingkai kesehatan mental sehingga menghasilkan 7 label utama: ‘*Awareness*’, ‘*Feelings and Problematization*’, ‘*Classification*’, ‘*Accessibility and Funding*’, ‘*Stigma*’, ‘*Service*’, dan ‘*Youth*’. Tahap kedua berfokus pada penyusunan dataset tweet Indonesia yang berjumlah 29.068 data, dengan menyaring tweet menggunakan kata kunci “*mental health*” dan “kesehatan mental”. Pada tahap ketiga, penelitian ini melakukan *preprocessing* data dan pelabelan manual terhadap 3.828 tweet yang dipilih secara acak, yang dipilih karena tidak praktisnya pelabelan seluruh data dan agar setiap data yang diambil acak memiliki peluang yang sama sehingga data tidak merujuk ke salah satu label saja. Akhirnya, pada tahap keempat melakukan eksperimen *topic modeling* dengan melibatkan 2 skenario yang menggunakan tiga algoritma berbeda. Evaluasi dilakukan dengan menggunakan *human interpretation* dan *coherence score*. Hasil eksperimen *topic modeling* menunjukkan bahwa algoritma LDA dan LSI lebih efektif daripada HDP dalam menghasilkan kata kunci yang relevan dengan setiap label utama. LDA menunjukkan peningkatan koherensi seiring bertambahnya cluster hingga batas tertentu. LSI menunjukkan kinerja koherensi yang beragam tergantung jumlah cluster, sedangkan HDP memiliki koherensi yang relatif stabil dari 3 hingga 7 cluster.

Kata Kunci: Kesehatan Mental, Pemodelan Topik, LDA, LSI, HDP

ABSTRACT

This research aims to implement topic modeling algorithms on mental health text tweet data, reviewing the literature on this topic reveals that previous research has a narrow focus, causing many issues to be overlooked. This paper addresses these issues by applying topic modeling methods across various mental health frames through several stages. First, the paper examines various mental health frames resulting in 7 main labels: 'Awareness', 'Feelings and Problematization', 'Classification', 'Accessibility and Funding', 'Stigma', 'Service', and 'Youth'. The second stage focused on compiling a dataset of 29,068 Indonesian tweets, by filtering tweets using the keywords “mental health” and “mental health”. In the third stage, this research conducted data preprocessing and manual labeling of 3,828 randomly selected tweets, which were chosen due to the impracticality of labeling the entire data and so that each randomly drawn data has an equal chance of not referring to any one label. Finally, in the fourth stage, topic modeling experiments were conducted involving two scenarios using three different algorithms. Evaluation was done using human interpretation and coherence score. The results of the topic modeling experiments show that the LDA and LSI algorithms are more effective than HDP in generating keywords relevant to each main label. LDA shows an increase in coherence as the number of clusters increases up to a certain limit. LSI shows mixed coherence performance depending on the number of clusters, while HDP has relatively stable coherence from 3 to 7 clusters.


Keywords: Mental Health, Topic Modeling, LDA, LSI, HDP

LEMBAR PERSEMBAHAN

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT, atas segala rahmat, hidayah, dan karunia-Nya yang selalu menyertai langkah saya dalam menyelesaikan Tugas Akhir ini.
2. Ibu Suliati dan Bapak Joni Effendy selaku orang tua saya yang membimbing saya menjadi pribadi yang baik hingga saat ini.
3. Saudara dan teman-teman saya seperjuangan yang menemani dan memberikan ilmu yang bermanfaat.
4. Bapak Setio Basuki, M.T., Ph.D selaku dosen pembimbing yang telah membimbing dalam melaksanakan pengerjaan tugas akhir.
5. Bapak Galih Wasis Wicaksono, S.Kom., M.Cs., selaku Ketua Jurusan Teknik Informatika Universitas Muhammadiyah Malang.
6. Seluruh Dosen beserta Staff Program Studi Informatika Universitas Muhammadiyah Malang yang telah memberikan ilmu dan wawasan kepada penulis selama melaksanakan studi.
7. Saya pribadi yang telah berhasil menyelesaikan studi.

Malang, 24 Agustus 2024



Nico Ardia Effendy

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT. Atas limpahan rahmat dan hidayah-NYA sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul:

“ IMPLEMENTASI ALGORITMA TOPIC MODELING PADA DATA CUITAN TERKAIT KESEHATAN MENTAL PADA PLATFORM MEDIA SOSIAL TWITTER (X) BERBAHASA INDONESIA”

Di dalam tulisan ini disajikan pokok-pokok bahasan yang meliputi latar belakang, metode penelitian, serta hasil dan pembahasan yang disimpulkan berdasarkan proses yang telah dilalui oleh peneliti.

Peneliti menyadari sepenuhnya bahwa dalam penulisan tugas akhir ini masih banyak kekurangan dan keterbatasan. Oleh karena itu, peneliti mengharapkan saran yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 24 Agustus 2024



Nico Ardia Effendy

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN.....	iv
ABSTRAK.....	iv
<i>ABSTRACT</i>	vi
LEMBAR PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
BAB II.....	6
TINJAUAN PUSTAKA	6
2.1 Pustaka Terdahulu	6
2.2 Definisi Kesehatan Mental	7
2.3 X (Twitter) dan Kesehatan Mental	7
2.4 Label Kesehatan Mental	7
2.5 Latent Dirichlet Allocation (LDA)	8
2.6 Latent Semantic Indexing (LSI)	9
2.7 Hierarchical Dirichlet Process (HDP).....	10
BAB III	11
METODOLOGI PENELITIAN.....	11
3.1 Dataset	11
3.2 Pelabelan Data Twitter	11
3.3 Pre-Processing Data Twitter	13
3.4 Skenario <i>Topic Modeling</i>	14
3.5 Evaluasi	14
BAB IV HASIL DAN PEMBAHASAN	16
4.1 Preprocessing dan Pelabelan Data Twitter	16
4.2 Hasil Eksperimen pada skenario pertama.....	16

4.3 Hasil Eksperimen pada skenario kedua	19
BAB V	24
KESIMPULAN.....	24
5.1 Kesimpulan.....	24
5.2 Saran	24
DAFTAR PUSTAKA	25



DAFTAR GAMBAR

Gambar 1. Representasi Grafis LDA	9
Gambar 2. Arsitektur Sistem Topic Modeling Terkait Kesehatan Mental pada Platform X (Twitter) Berbahasa Indonesia	11
Gambar 3. Distribusi semua kumpulan data sampel berlabel	13
Gambar 4. Distribusi kata kunci terbagi menjadi terkait, tidak terkait, dan netral pada LDA, LSI, dan HDP Schema	18
Gambar 5. Perbandingan nilai koheren untuk 3 cluster yang dihasilkan oleh algoritma LDA, LSI dan HDP menggunakan data keseluruhan teks cuitan	21
Gambar 6. Perbandingan nilai koheren untuk 5 cluster yang dihasilkan oleh algoritma LDA, LSI dan HDP menggunakan data keseluruhan teks cuitan	21
Gambar 7. Perbandingan nilai koheren untuk 7 cluster yang dihasilkan oleh algoritma LDA, LSI dan HDP menggunakan data keseluruhan teks cuitan	22



DAFTAR TABEL

Tabel 1. Pustaka Terdahulu.....	6
Tabel 2. Daftar top 10 kata kunci di setiap 7 label utama Kesehatan mental	18
Tabel 3. Daftar Top 10 Kata Kunci 7 cluster dengan keseluruhan data teks cuitan	22



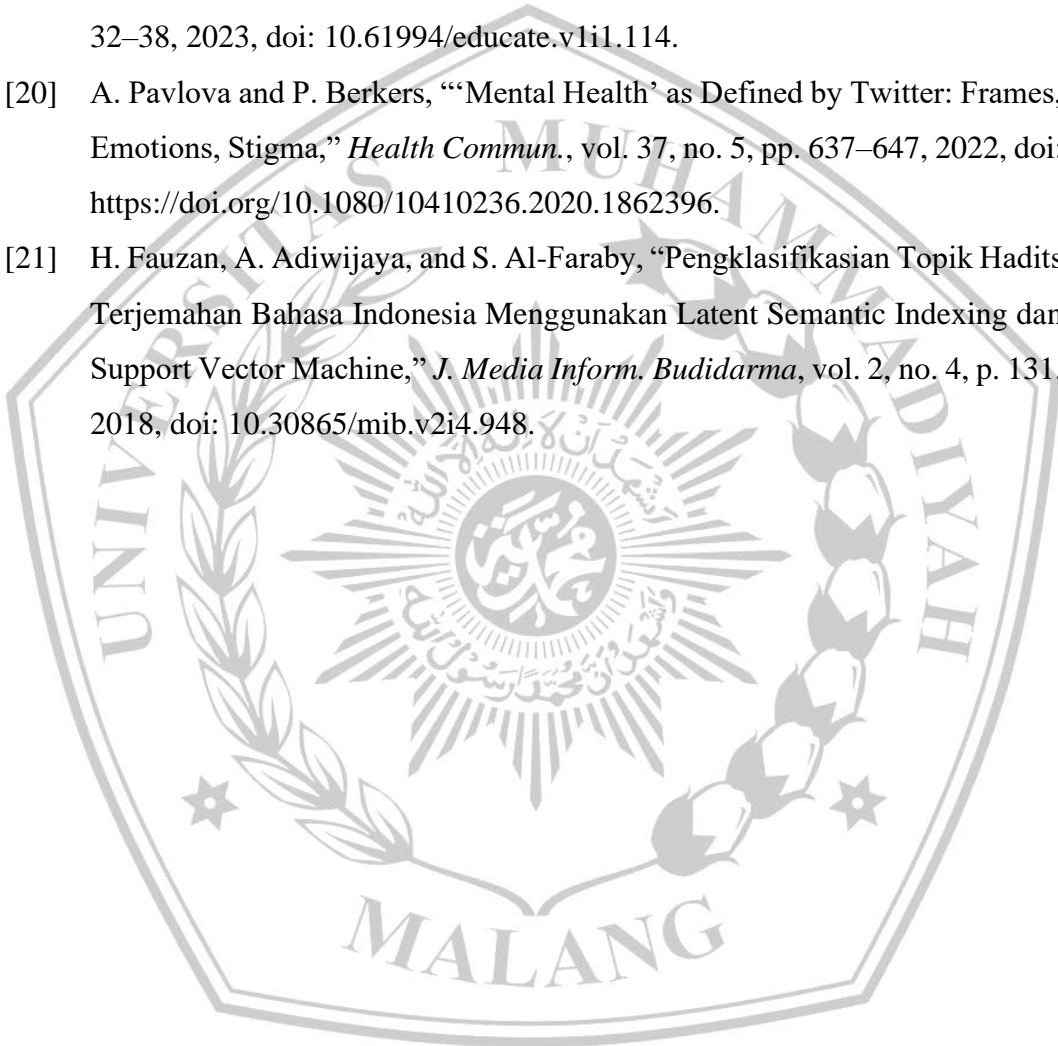
DAFTAR PUSTAKA

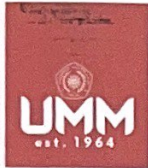
- [1] World Health Organization, *Mental Health Atlas 2020*. 2020. [Online]. Available: <https://www.who.int/publications/i/item/9789240036703>
- [2] World Health Organization, “Mental Health and COVID-19 : Early evidence of the pandemic ’ s impact,” *Sci. Br.*, vol. 2, no. March, pp. 1–11, 2022, [Online]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1
- [3] Ellyana Dwi Farisandy, Azzahra Asihputri, and Jennifer Shalom Pontoh, “Peningkatan Pengetahuan Dan Kesadaran Masyarakat Mengenai Kesehatan Mental,” *Disem. J. Pengabd. Kpd. Masy.*, vol. 5, no. 1, pp. 81–90, 2023, doi: <https://doi.org/10.33830/diseminasiabdimas.v5i1.5037>.
- [4] I. A. Ridlo, “Jurnal Psikologi dan Kesehatan Mental Pandemi COVID-19 dan Tantangan Kebijakan Kesehatan Mental di Indonesia,” *Dep. Adm. dan Kebijak. Kesehatan, Fak. Kesehat. Masy. Univ. Airlangga*, pp. 155–164, 2020, doi: [10.20473/jpkm.v5i12020.155-164](https://doi.org/10.20473/jpkm.v5i12020.155-164).
- [5] A. Thieme, D. Belgrave, and G. Doherty, “Machine Learning in Mental Health: A systematic review of the HCI literature to support the development of effective and implementable ML Systems,” *ACM Trans. Comput. Interact.*, vol. 27, no. 5, 2020, doi: <https://doi.org/10.1145/3398069>.
- [6] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, “Machine learning in mental health: A scoping review of methods and applications,” *Psychol. Med.*, vol. 49, no. 9, pp. 1426–1448, 2019, doi: <https://doi.org/10.1017/S0033291719000151>.
- [7] N. H. Di Cara, V. Maggio, O. S. P. Davis, and C. M. A. Haworth, “Methodologies for Monitoring Mental Health on Twitter: Systematic Review,” *J. Med. Internet Res.*, vol. 25, pp. 1–21, 2023, doi: <https://doi.org/10.2196/2F42734>.
- [8] L. Liu and B. K. P. Woo, “Twitter as a mental health support system for students and professionals in the medical field,” *JMIR Med. Educ.*, vol. 7, no. 1, 2021, doi: <https://doi.org/10.2196/17598>.
- [9] A. Le Glaz *et al.*, “Machine learning and natural language processing in mental health: Systematic review,” *J. Med. Internet Res.*, vol. 23, no. 5, 2021,

doi: <https://doi.org/10.2196/15708>.

- [10] I. Dwijayanti *et al.*, “Tweets Classification of Mental Health Disorder in Indonesia Using LDA and Cosine Similarity,” vol. 21, no. 1, pp. 41–54, 2024, doi: 10.31515/telematika.v21i1.10725.
- [11] P. Damayanti, D. Purwitasari, and N. Suciati, “Eliminasi Non-Topic Menggunakan Pemodelan Topik untuk Peringkasan Otomatis Data Tweet dengan Konteks Covid-19,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 199, 2021, doi: 10.25126/jtiik.0814324.
- [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. Transitions Proc.*, vol. 3, no. 1, pp. 298–304, 2022, doi: <https://doi.org/10.1016/j.gltip.2022.03.015>.
- [13] B. Ogunleye, T. Maswera, L. Hirsch, J. Gaudoin, and T. Brunson, “Comparison of Topic Modelling Approaches in the Banking Context,” *Appl. Sci. 2023, Vol. 13, Page 797*, vol. 13, no. 2, p. 797, Jan. 2023, doi: <https://doi.org/10.3390/app13020797>.
- [14] S. Koltcov, V. Ignatenko, M. Terpilovskii, and P. Rosso, “Analysis and tuning of hierarchical topic models based on Renyi entropy approach,” *PeerJ Comput. Sci.*, vol. 7, pp. 1–35, 2021, doi: <https://doi.org/10.7717/peerj-cs.608/supp-1>.
- [15] K. R. A. P. Santoso, A. Husna, N. W. Putri, and N. A. Rakhmawati, “Analisis Topik Tagar Covidindonesia pada Instagram Menggunakan Latent Dirichlet Allocation,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.14421/jiska.2022.7.1.1-9.
- [16] D. Zakeshia Tiara Kannitha and P. Kartikasari, “PEMODELAN TOPIK PADA KELUHAN PELANGGAN MENGGUNAKAN ALGORITMA LATENT DIRICHLET ALLOCATION DALAM MEDIA SOSIAL TWITTER,” vol. 11, no. 2, pp. 266–277, 2022, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [17] Khairul Hudha Nasution, Widodo, and Bambang Prasetya Adhi, “Sistem Deteksi Topik Politik Pada Twitter Menggunakan Algoritma Latent Dirichlet Allocation,” *PINTER J. Pendidik. Tek. Inform. dan Komput.*, vol.

- 5, no. 1, pp. 76–83, 2021, doi: 10.21009/pinter.5.1.10.
- [18] A. Rosmalina and T. Khaerunnisa, “Penggunaan Media Sosial dalam Kesehatan Mental Remaja,” *Prophet. Prof. Empathy, Islam. Couns. J.*, vol. 4, no. 1, p. 49, 2021, doi: 10.24235/prophetic.v4i1.8755.
- [19] Aisyah Fitriah, Dzaky Juliansyah, Umi Salamah, M Anugrah Utama, Opie Karunia Falah, and Aseh Miati, “Pengaruh Media Sosial Terhadap Kesehatan Mental Pada Remaja,” *Educ. J. Educ. Learn.*, vol. 1, no. 1, pp. 32–38, 2023, doi: 10.61994/educate.v1i1.114.
- [20] A. Pavlova and P. Berkers, “‘Mental Health’ as Defined by Twitter: Frames, Emotions, Stigma,” *Health Commun.*, vol. 37, no. 5, pp. 637–647, 2022, doi: <https://doi.org/10.1080/10410236.2020.1862396>.
- [21] H. Fauzan, A. Adiwijaya, and S. Al-Faraby, “Pengklasifikasian Topik Hadits Terjemahan Bahasa Indonesia Menggunakan Latent Semantic Indexing dan Support Vector Machine,” *J. Media Inform. Budidarma*, vol. 2, no. 4, p. 131, 2018, doi: 10.30865/mib.v2i4.948.





UNIVERSITAS
MUHAMMADIYAH
MALANG



FAKULTAS TEKNIK

INFORMATIKA

informatika.umm.ac.id | informatika@umm.ac.id

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Nico Ardia Effendy
NIM : 202010370311467
Judul TA : Implementasi Algoritma Topic Modeling pada Data Cuitan
 Terkait Kesehatan Mental pada Platform Media Sosial Twitter (X)
 Berbahasa Indonesia

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	9%
2.	Bab 2 – Daftar Pustaka	25 %	18%
3.	Bab 3 – Analisis dan Perancangan	25 %	0%
4.	Bab 4 – Implementasi dan Pengujian	15 %	5%
5.	Bab 5 – Kesimpulan dan Saran	5 %	0%
6.	Makalah Tugas Akhir	20%	12%

*) Hasil cek plagiarism 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 Sutami No. 188 Malang, Jawa Timur
 P. +62 341 551 149 (Hunting)
 F. +62 341 582 060

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