

**PENGELOMPOKAN TOPIK OPINI MASYARAKAT TERHADAP
PROGRAM SEKOLAH RAKYAT MENGGUNAKAN *LATENT*
*DIRICHLET ALLOCATION***

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh
AFZENI FITRI HANIK
09040222049

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2026

LEMBAR PERSETUJUAN PEMBIMBING

skripsi oleh

Nama : Afzeni Fitri Hanik
NIM : 09040222049
Judul skripsi : Pengelompokan Topik Opini Masyarakat Terhadap Program Sekolah Rakyat Menggunakan *Latent Dirichlet Allocation*

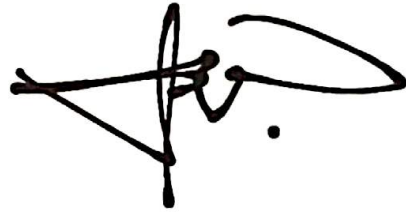
telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



Nuri Saidah Ulinuha, M. Kom
NIP. 199011022014032004

Pembimbing II



Dian Yuliati, M.Si
NIP. 198707142020122015

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya



Dr. Yuniar Farida, M.T
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : Afzeni Fitri Hanik
NIM : 09040222049
Judul Skripsi : Pengelompokan Topik Opini Masyarakat Terhadap Program Sekolah Rakyat Menggunakan *Latent Dirichlet Allocation*

Telah dipertahankan di depan Tim Penguji
pada tanggal 05 Januari 2026

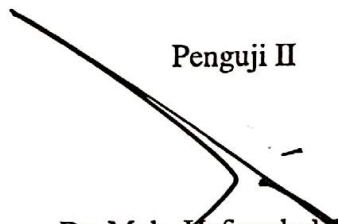
Mengesahkan,
Tim Penguji

Penguji I



Aris Fanani, M.Kom
NIP. 198701272014031002

Penguji II



Dr. Moh. Hafiyusholek, M.Si., M.PMat.
NIP. 198002042014031001

Penguji III



Nurissaidah Ulinnuha, M. Kom
NIP. 199011022014032004

Penguji IV



Dian Yuliati, M.Si
NIP. 198707142020122015

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya



Dr. Saepul Hamdani, M.Pd
NIP. 196507312000031002

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : Afzeni Fitri Hanik

NIM : 09040222049

Program Studi : Matematika

Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "Pengelompokan Topik Opini Masyarakat Terhadap Program Sekolah Rakyat Menggunakan *Latent Dirichlet Allocation*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 05 Januari 2026

Yang menyatakan,




Afzeni Fitri Hanik

NIM. 09040222049

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Afzani Fitri Hanik
NIM : 09040222049
Fakultas/Jurusan : SAINTEK / MATEMATIKA
E-mail address : afzenfitr@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

PENGELOMPOKAN TOPIK OPINI MASYARAKAT TERHADAP
PROGRAM SEKOLAH RAKYAT MENGGUNAKAN LATENT DIRICHLET
ALLOCATION

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 05 Januari 2026

Penulis



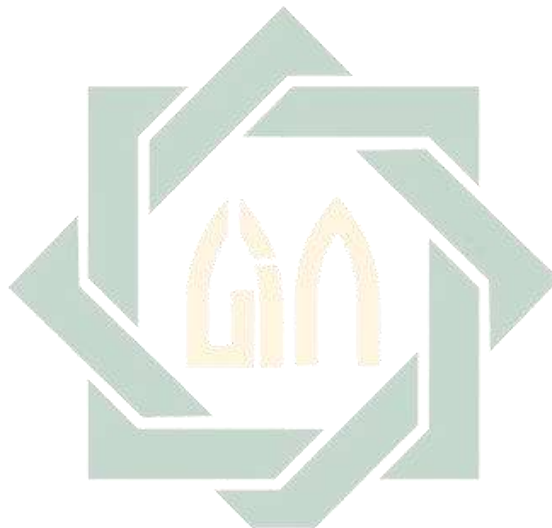
(Afzani Fitri Hanik)
nama terang dan tanda tangan

DAFTAR ISI

MOTTO	1
PERSEMBAHAN	2
KATA PENGANTAR	4
DAFTAR ISI	6
DAFTAR TABEL	9
DAFTAR GAMBAR	11
ABSTRAK	12
ABSTRACT	13
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	7
1.3. Tujuan Penelitian	7
1.4. Manfaat Penelitian	7
1.5. Batasan Masalah	8
1.6. Sistematika Penulisan	9
II TINJAUAN PUSTAKA	10
2.1. Sekolah Rakyat	10
2.2. YouTube dan TikTok	12
2.3. Pengumpulan Data Media Sosial	13
2.3.1. Web Scraping	13
2.3.2. <i>Application Programming Interface(API)</i>	13
2.3.3. APIFY	14
2.4. Text Mining	14
2.5. Pemodelan Topik	15
2.6. Text Preprocessing	15
2.6.1. Remove Punctuation	15
2.6.2. Case Folding	16

2.6.3. Tokenizing	16
2.6.4. Normalisasi	17
2.6.5. Stopword	17
2.6.6. Stemming	18
2.7. N-Gram	18
2.8. Representasi Bag of Words	19
2.9. <i>Latent Dirichlet Allocation (LDA)</i>	20
2.10. <i>Collapsed Gibbs Sampling</i>	24
2.11. Topik Coherence	26
2.12. Wordcloud	29
2.13. Integrasi Keislaman	31
III METODE PENELITIAN	34
3.1. Jenis Penelitian	34
3.2. Sumber Data	34
3.3. Tahapan Penelitian	35
3.4. Skenario Uji Coba Parameter	39
IV HASIL DAN PEMBAHASAN	40
4.1. Pengambilan Data	40
4.2. Preprocessing	46
4.2.1. Remove Punctuation	46
4.2.2. Case Folding	47
4.2.3. Tokenizing	48
4.2.4. Normalisasi	49
4.2.5. Stemming	50
4.2.6. Stopword	51
4.2.7. N Gram	52
4.3. Representasi Bag Of Words	55
4.4. Pemodelan Topik LDA dengan <i>Gibbs Sampling</i>	56
4.5. Uji Coba Parameter	60
4.6. Visualisasi dan Analisis Pemodelan Topik	63
4.6.1. Hasil Pengelompokan Data (Lampiran)	73
4.6.2. Visualisasi dan Interpretasi Topik ke-1 LDA	74
4.6.3. Visualisasi dan Interpretasi Topik 2 LDA	76
4.6.4. Visualisasi dan Interpretasi Topik 3 LDA	78

4.6.5. Visualisasi dan Interpretasi Topik 4 LDA	80
4.6.6. Visualisasi dan Interpretasi Topik 5 LDA	82
4.6.7. Visualisasi dan Interpretasi Topik 6 LDA	84
4.6.8. Visualisasi dan Interpretasi Topik ke-7 LDA	86
4.7. Integrasi Keislaman	88
V PENUTUP	93
5.1. Kesimpulan	93
5.2. Saran	94
DAFTAR PUSTAKA	95



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

4.1 Daftar Video YouTube Sumber Data Penelitian	45
4.2 Daftar Video TikTok Sumber Data Penelitian	46
4.3 Proses Remove Punctuation	47
4.4 Proses Case Folding	48
4.5 Proses Tokenizing	49
4.6 Proses Normalisasi	50
4.7 Proses Stemming	51
4.8 Proses Stopword	52
4.9 Daftar Bigram yang Terbentuk	53
4.10 Daftar Trigram yang Terbentuk	53
4.11 Hasil Bigram	54
4.12 Hasil Trigram	55
4.13 Representasi Data Teks Menggunakan Metode Bag of Words	56
4.14 Inisialisasi Topik Awal Setiap Token	57
4.15 Jumlah kemunculan kata pada setiap topik	58
4.16 Jumlah Kemunculan Kata pada Setiap Topik	59
4.17 Jumlah kemunculan topik pada seluruh dokumen	59
4.18 Nilai Koherensi pada Berbagai Kombinasi Parameter	61
4.19 Koherensi tiap topik pada model terbaik	64
4.20 Hasil Pemodelan Topik LDA	65
4.21 Hasil Preprocessing	66
4.22 Inisialisasi Untuk 7 Topik Awal Setiap Token	66
4.23 Contoh tabel topik dan kata	67
4.24 Jumlah nk per topik	67
4.25 Nilai Kemunculan Topik pada Setiap Dokumen	67
4.26 Contoh Perhitungan Probabilitas Kata <i>program</i> pada Dokumen ke-1	68

4.27 Contoh Perhitungan Probabilitas Kata <i>program</i> pada	
Dokumen ke-3	69
4.28 Perhitungan Probabilitas Kata <i>kata</i> pada Dokumen ke-1	70
4.29 Contoh Perhitungan Probabilitas Kata <i>anak_abah</i> pada	
Dokumen ke-1	70
4.30 Contoh Perhitungan Probabilitas Kata <i>didik</i> pada Dokumen	
ke-1	71
4.31 Contoh Perhitungan Probabilitas Kata <i>penting</i> pada	
Dokumen ke-1	71
4.32 Contoh Perhitungan Probabilitas Kata <i>makan_gratis</i> pada	
Dokumen ke-1	72
4.33 Contoh Perhitungan Probabilitas Kata <i>bapak_prabowo</i> pada	
Dokumen ke-1	72
4.34 Distribusi Topik Kata pada Dokumen ke-1	73
4.35 Hasil LDA Pengelompokan Topik	74
4.36 Distribusi Kata Topik ke-1	76
4.37 Contoh Komentar Topik 1	76
4.38 Distribusi Kata Topik ke-2	78
4.39 Contoh Komentar Topik 2	78
4.40 Distribusi Kata Topik ke-3	80
4.41 Contoh Komentar Topik 3	80
4.42 Distribusi Kata Topik ke-4	82
4.43 Contoh Komentar Topik 4	82
4.44 Distribusi Kata Topik ke-5	83
4.45 Contoh Komentar Topik 5	84
4.46 Distribusi Kata Topik ke-6	85
4.47 Contoh Komentar Topik 6	86
4.48 Distribusi Kata pada Topik ke-7	87
4.49 Contoh Komentar Topik 7	88

DAFTAR GAMBAR

2.1 Model Grafis LDA	23
2.2 Alur Topic Coherence	27
2.3 Ilustrasi Wordcloud	31
3.1 Flowchart Penelitian	35
3.2 Flowchart LDA Gibbs Sampling	37
4.1 Data Komentar YouTube	41
4.2 Komentar YouTube bulan Juli	41
4.3 Tampilan Awal Apify	42
4.4 TikTok Comment Scrapper	42
4.5 TikTok video url	43
4.6 Data Komentar TikTok	43
4.7 Komentar TikTok bulan Agustus	44
4.8 Rata-rata Nilai Koherensi Tiap Parameter	62
4.9 wordcloud topik 1	75
4.10 wordcloud topik 2	77
4.11 wordcloud topik 3	79
4.12 wordcloud topik 4	81
4.13 wordcloud topik 5	83
4.14 wordcloud topik 6	85
4.15 wordcloud topik 7	87

ABSTRAK

Pengelompokan Topik Opini Masyarakat Terhadap Program Sekolah Rakyat Menggunakan *Latent Dirichlet Allocation*

Pemerataan akses pendidikan merupakan tantangan utama dalam pembangunan sumber daya manusia di Indonesia, khususnya bagi masyarakat kurang mampu. Program Sekolah Rakyat hadir sebagai upaya pemerintah untuk mengurangi ketimpangan pendidikan dan menekan angka putus sekolah. Sejak diumumkan, program ini memunculkan beragam tanggapan masyarakat yang disampaikan melalui media sosial, terutama YouTube dan TikTok. Penelitian ini bertujuan untuk mengevaluasi efektivitas metode *Latent Dirichlet Allocation* (LDA) dalam mengelompokan topik pada data komentar YouTube dan TikTok, serta memberikan gambaran mengenai opini masyarakat terhadap Program Sekolah Rakyat berdasarkan hasil pemodelan topik. Data penelitian berupa 6.839 komentar yang dikumpulkan pada periode Maret hingga Oktober 2025. Evaluasi model dilakukan menggunakan nilai koherensi untuk menentukan jumlah topik dan parameter terbaik. Hasil pengujian menunjukkan bahwa model terbaik diperoleh pada 7 topik dengan alpha 0.5, beta 0.05, dan 50 iterasi, menghasilkan nilai koherensi tertinggi sebesar 0,648. Topik yang terbentuk mencerminkan dukungan, harapan, dan kritik masyarakat terhadap pelaksanaan program.

Kata kunci: LDA, Media Sosial, Pengelompokan Topik, Sekolah Rakyat

UIN SUNAN AMPEL
S U R A B A Y A

ABSTRACT

Topic Clustering of Public Opinion on the Sekolah Rakyat Program Using Latent Dirichlet Allocation

Equitable access to education is a major challenge in human resource development in Indonesia, particularly for underprivileged communities. The People's School Program (Sekolah Rakyat) is a government effort to reduce educational inequality and reduce the dropout rate. Since its announcement, the program has generated diverse public responses through social media, particularly YouTube and TikTok. This study aims to evaluate the effectiveness of the Latent Dirichlet Allocation (LDA) method in grouping topics in YouTube and TikTok comment data and to provide an overview of public opinion on the People's School Program based on topic modeling results. The research data consisted of 6,839 comments collected from March to October 2025. Model evaluation was conducted using coherence values to determine the best number of topics and parameters. Test results showed that the best model was obtained with 7 topics with an alpha of 0.5, beta of 0.05, and 50 iterations, resulting in the highest coherence value of 0.648. The resulting topics reflect public support, expectations, and criticism of the program's implementation.

Keywords: LDA, Social Media, Topic Clustering, People's School



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR PUSTAKA

- Al Fauziah, S. A., Budhiharti, W., and Rahman, K. A. (2025). Pengaruh postingan instagram@ lambe_turah terhadap kesiapan untuk menikah (studi kasus perceraian baim wong dan paula verhoeven). *Jurnal Multidisiplin Dehasen (MUDE)*, 4(3):607–616.
- Aletras, N. and Stevenson, M. (2013). Evaluating topic coherence using distributional semantics. In *Proceedings of the 10th international conference on computational semantics (IWCS 2013)–Long Papers*, pages 13–22.
- Alfanzar, A. I., Khalid, K., and Rozas, I. S. (2020). Topic modelling skripsi menggunakan metode latent dirichlet allocation. *JSiI (Jurnal Sistem Informasi)*, 7(1):7–13.
- Amalia, J. (2022). Membangun slang dictionary untuk normalisasi teks menggunakan pre-trained fasttext model. *JSR: Jaringan Sistem Informasi Robotik*, 6(2):250–256.
- Ariansyah, A. and Indahyanti, U. (2024). Fitur ekstraksi pada pemodelan topik menggunakan metode latent dirichlet allocation pada peristiwa kebocoran data. *Indonesian Journal of Applied Technology*, 1(2):24–24.
- Asgar, H., Hartono, B., et al. (2022). Rancang bangun rest api aplikasi weshare sebagai upaya mempermudah pelayanan donasi kemanusiaan. *Jurnal Informatika Teknologi dan Sains (Jinteks)*, 4(1):8–14.
- Aziz, R., Fahrudin, T. M., and Saputra, W. S. J. (2024). Analisis sentimen kepuasan pengguna oyo diplaystore dengan multinomial naive bayes dan chi-square. *Jurnal Fasilkom*, 14(1):166–175.
- Bellaouar, S., Bellaouar, M. M., and Ghada, I. E. (2021). Topic modeling: Comparison of lsa and lda on scientific publications. In *Proceedings of the 2021 4th International Conference on Data Storage and Data Engineering*, pages 59–64.

- Cahyono, N. (2024). *Fundamental Topic Modelling: Memahami Tren & Topik dalam Text Mining*. MEGA PRESS NUSANTARA.
- Calista, C. P., Putra, F. A. I., Maulita, F., Zahra, M., and Syakirah, N. A. (2025). Kesenjangan pendidikan antara wilayah pedesaan dan perkotaan: Analisis kebijakan dan praktik di Indonesia. *Integrative Perspectives of Social and Science Journal*, 2(03 Juni):3965–3978.
- Dewi, N. K. (2023). Identifikasi berita hoax dengan menerapkan algoritma text mining. *Journal of Informatics, Electrical and Electronics Engineering*, 2(3):65–74.
- Fatah, M., Yohanes Ngamal, S., and Setiawati, E. (2024). Kemajuan tik, digitalisasi membuka jalan bagi pertumbuhan ekonomi digital Indonesia. *Jurnal Manajemen & Bisnis*, 16(II).
- Fazri, M. and Voutama, A. (2025). Analisis sentimen publik terhadap danantara di media sosial x menggunakan nlp dan pembelajaran mesin. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, 9(1):197–206.
- Hakim, L., Peryanto, A., Susanto, D., and Widodo, Y. F. (2025). Meninjau peranan text mining sebagai alat strategis dalam industri kreatif melalui sajian webinar. *Jurnal Pengabdian Masyarakat-PIMAS*, 4(3):225–233.
- Hamdan, H. and Mahmuddin, M. (2021). Youtube sebagai media dakwah. *Palita: Journal of Social Religion Research*, 6(1):63–80.
- Harahap, S. and Arisman, A. (2024). Maqashid al-syari'ah berdasarkan kemaslahatan yang dipelihara dalam hukum Islam. *HUKUMAH: Jurnal Hukum Islam*, 6(2):92–106.
- Hayatulnupus, S. M. (2025). Transformasi komunikasi masyarakat modern melalui tiktok: Analisis partisipatif generasi z. *Bashirah: Jurnal Komunikasi dan Penyiaran Islam*, 6(1):94–106.
- Hsu-Hua, L. and Minh, T. (2023). Topic modelling and sentiment analysis on youtube sustainable fashion comments. *Journal of New Media*, 5(1):65.

- Hudaidah, H. and Karwana, M. A. P. (2022). Pendidikan di indonesia masa pendudukan jepang. *Danadyaksa Historica*, 1(2):97–104.
- Huwaida, S. F., Kusumawati, R., and Isnaini, B. (2024). Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode naïve bayes. *Jambura Journal of Informatics*, 6(1):26–39.
- Ichida, N. A. F. (2024). Optimasi algoritma simplified lesk dengan spacy untuk word sense disambiguation pada kalimat bahasa inggris. *Jurnal Khatulistiwa Informatika*, 12(1):13–18.
- Indrayanto, C. G., Ratnawati, D. E., and Rahayudi, B. (2023). Analisis sentimen data ulasan pengguna aplikasi mypertamina di indonesia pada google play store menggunakan metode random forest. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(3):1131–1139.
- Jin, Y. (2017). Development of word cloud generator software based on python. *Procedia engineering*, 174:788–792.
- Jurafsky, D. and Martin, J. H. (2023). *Speech and Language Processing*. Pearson, Hoboken, NJ, 3 edition.
- Kannitha, D. Z. T., Mustafid, M., and Kartikasari, P. (2022). Pemodelan topik pada keluhan pelanggan menggunakan algoritma latent dirichlet allocation dalam media sosial twitter. *Jurnal Gaussian*, 11(2):266–277.
- Kompas.com (2025). Survei apjii: 5 medsos terpopuler di indonesia 2025, tiktok juara. Diakses pada 23 Oktober 2025.
- Kristanto, T., Prayudi, M. N., and Rochmah, W. Y. (2025). Analisis penggunaan media sosial untuk meningkatkan brand awareness di youtube. *Jurnal Ilmiah Bisnis Digital*, 1(2):67–76.
- Kustiyahningsih, Y. and Permana, Y. (2024). Penggunaan latent dirichlet allocation (lda) dan support-vector machine (svm) untuk menganalisis sentimen berdasarkan aspek dalam ulasan aplikasi edlink. *Teknika*, 13(1):127–136.

liputan6.com (2025). Depan anggota dpd, wamensos tegaskan jumlah sekolah rakyat akan terus bertambah. Diakses pada 23 Oktober 2025.

Mauliza, R. N. and Sipayung, Y. R. (2024). Penerapan text mining dalam menganalisis pendapat masyarakat terhadap pemilu 2024 pada media sosial x menggunakan metode naive bayes. *Technomedia Journal*, 9(1):1–16.

MetroTVNews.com (2025). 13 sekolah rakyat di jabar menampung 1.390 siswa. Diakses pada 23 Oktober 2025.

Misrun, C. A., Haerani, E., Fikry, M., and Budianita, E. (2023). Analisis sentimen komentar youtube terhadap anies baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier. *jurnal coscitech (computer science and information technology)*, 4(1):207–215.

Mohammed, S. H. and Al-augby, S. (2020). Lsa & lda topic modeling classification: Comparison study on e-books. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(1):353–362.

MPR (2025). Mpr edisi no 7 2025. Diakses pada 23 Oktober 2025.

Muluk, M. F. A. M. (2024). *ANALISIS SENTIMEN PENGGUNA TIKTOK TERHADAP PRODUK FASHION DENGAN METODE NAÏVE BAYES*. PhD thesis, UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA.

Muna, Z. N. (2024). Penerapan pemodelan topik komentar melalui media sosial twitter menggunakan latent dirichlet allocation (studi kasus: Pemerintah kota malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(7).

Novriyanto, Y., Apriyana, S. B., and Komariyah, S. (2022). Pengaruh pengaruh kebijakan politik etis terhadap perkembangan pendidikan di indonesia: awal kebijakan politik etis terhadap pendidikan, sistem pendidikan di zaman belanda, lembaga pendidikan belanda. *Jurnal Ilmu Pendidikan Dan Sosial*, 1(1):88–94.

Nuradilla, S., Kamila, S. A., Zahra, L., Suhaeni, C., and Sartono, B. (2025). Pemodelan topik pada komentar youtube arra: Komparasi lda dan k-

means menggunakan fitur leksikal dan semantik. *Jurnal Informatika: Jurnal Pengembangan IT*, 10(3):694–707.

Nurhaliza, A. F. (2024). Penerapan pemodelan topik menggunakan metode latent dirichlet allocation terhadap pembahasan pemilu indonesia tahun 2024 di twitter. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(7).

Nurrosidha, A., Fahriyyah, J., Fazran, R. F., and Alpiqri, A. (2022). Kaidah keyakinan atas keraguan dan implementasinya dalam urusan pemerintahan. *NALAR FIQH: Jurnal Hukum Islam*, 13(1):1–12.

Nursaadah, S. and Cahyana, R. (2025). Pemodelan analisis tren topik penelitian sistem informasi menggunakan latent dirichlet allocation. *Jurnal Algoritma*, 22(2):1263–1274.

Oktavianaa, A. K. N., ERa, N. A. S., Mahendraa, I. B. M., Astawaa, I. G. S., Wibawaa, I. G. A., and Mogia, I. K. A. (2022). Pemodelan topik artikel berita menggunakan structural topic model dan latent dirichlet allocation. *Jurnal Elektronik Ilmu Komputer Udayana p-ISSN*, 2301:5373.

Pardede, D. C. and Waskita, M. A. I. (2023). Analisis pemodelan topik untuk ulasan tentang peduli lingkungan. *Jurnal Ilmiah Informatika Komputer*, 28(1):17–26.

Purwitasari, D., Muflichah, A., Hasanah, N. A., Arifin, A. Z., et al. (2021). Pemodelan topik dengan lda untuk temu kembali informasi dalam rekomendasi tugas akhir. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(3):421–428.

Putra, K. T., Anggraini, S., Sutriani, L., Suraji, S., and Impron, A. (2025). Analisis sentimen masyarakat kalimantan tengah terhadap perkebunan kelapa sawit menggunakan tf-idf dan support vector machine. *Journal Scientific of Mandalika (JSM) e-ISSN 2745-5955— p-ISSN 2809-0543*, 6(5):1115–1123.

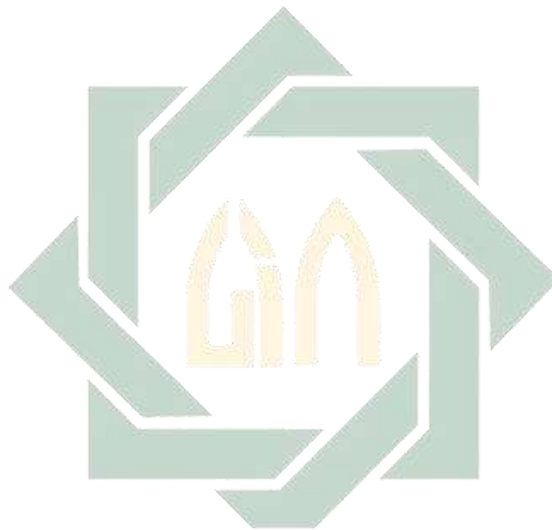
RI, D. (2025). Dpr ri 2025. Diakses pada 1 November 2025.

- Röder, M., Both, A., and Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408.
- Rosadi, Z. and Solichin, A. (2024). Topic modeling tugas akhir mahasiswa menggunakan metode latent dirichlet allocation dengan gibbs sampling. *Jurnal Ticom: Technology of Information and Communication*, 13(1):38–44.
- Rosalinda, G., Santoso, R., and Kartikasari, P. (2023). Pemodelan topik ulasan aplikasi netflix pada google play store menggunakan latent dirichlet allocation. *Jurnal Gaussian*, 11(4):554–561.
- Saputra, H. T., Damanik, A., Shaquille, M. M., Rusydi, M. A., Riziq, M. H., and Zulva, N. S. (2025). Analisis modelling pada reviews lazada indonesia menggunakan latent dirichlet allocation (lda) untuk optimalisasi strategi bisnis. *JEKIN-Jurnal Teknik Informatika*, 5(1):361–371.
- Sari, D. F., Kusjani, A., Kurniawati, D., and Setiawan, I. (2023). Pencarian data quick count pilpres dengan teknik web scraping. *Journal of Innovation Research and Knowledge*, 3(5):1025–1034.
- Sepriadi, H., Sekarsari, C. R., Iriany, A., Fernandes, A. A. R., et al. (2025). Kombinasi analisis bibliometrik dengan latent dirichlet allocation sebagai pemodelan topik cashless society. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(2):331–338.
- Setijohatmo, U. T., Rachmat, S., Susilawati, T., and Rahman, Y. (2020). Analisis metoda latent dirichlet allocation untuk klasifikasi dokumen laporan tugas akhir berdasarkan pemodelan topik. In *Prosiding Industrial Research Workshop and National Seminar*, volume 11, pages 402–408.
- Shohibuddin, M. and Maimun, M. (2025). Sekolah rakyat sebagai alternatif pendidikan di indonesia pada masa kolonial. *Journal of Linguistics and Social Studies*, 2(1):12–23.
- Statistik, B. P. (2024). Statistik pendidikan indonesia tahun 2024. Diakses pada 1 Oktober 2025.

- Suhaeni, C., Mualifah, L. N. A., and Wijayanto, H. (2025). Lda topic modeling analysis of public discourse on indonesia's free nutritious meals program (mbg). *IJID (International Journal on Informatics for Development)*, 14(1):587–600.
- Suharsono, J. P. and Nurahman, D. (2024). Pemanfaatan youtube sebagai media peningkatan pelayanan dan informasi. *Ganaya: Jurnal Ilmu Sosial dan Humaniora*, 7(1):298–304.
- Tiara, L., Syaputra, H., Cholil, W., and Mirza, A. H. (2021). Web scraper dan graphql api untuk data perguruan tinggi di indonesia berdasarkan website kementerian ristekdikti (studi kasus: Website kementerian ristekdikti). *Jurnal Nasional Ilmu Komputer*, 2(3):193–212.
- Tjen, J. (2025). Pengaruh jenis stopwords terhadap akurasi model multinomial naïve bayes dalam proses sentimen analisis. *Jurnal Buana Informatika*, 16(01):13–22.
- Wardhani, D., Astuti, R., and Saputra, D. D. (2024). Optimasi feature selection text mining: Stemming dan stopword untuk sentimen analisis aplikasi saturehat. *Innovative: Journal Of Social Science Research*, 4(1):7537–7548.
- Wear, P. N. et al. (2024). Analisis sentimen terhadap hasil pilpres 2024 pada aplikasi tiktok dan x menggunakan metode naïve bayes classifier. *JURNAL ILMIAH INFORMATIKA MULIA*, 1(2):15–22.
- Wibowo, J. A., Mawardi, V. C., and Sutrisno, T. (2024). Visualisasi word cloud hasil analisis sentimen berbasis fitur layanan aplikasi gojek dengan support vector machine. *Jurnal Serina Sains, Teknik dan Kedokteran*, 2(1):61–70.
- Yolanda, Y. M., Swanjaya, D., and Pamungkas, D. P. (2025). Sistem pencarian dongeng berbasis topik menggunakan latent dirichlet allocation (lda) pada aplikasi nusantara panji kediri. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 10(2):394–407.
- Zaza, D. A. U., Umar, E., and Sanga, F. E. O. (2024). Penerapan text mining dalam klasifikasi judul skripsi mahasiswa pada universitas stella maris

sumba menggunakan metode naïve bayes. *Journal Of Informatics And
Business*, 2(3):327–337.

Zuriatin, Z. (2022). Perkembangan pendidikan dan pengajaran masa
pendudukan jepang tahun 1942-1945. *Pendidikas: Jurnal Pendidikan
Dasar*, 3(01):24–30.



UIN SUNAN AMPEL
S U R A B A Y A