INFOTECH journal
ISSN : 2460-1861 (Print), 2615-4250 (Online)
Vol. 12 No. 1 2026

PEMODELAN TOPIK BERITA NASIONAL INDONESIA MENGGUNAKAN *LATENT DIRICHLET ALLOCATION*

Fajar Maula Hidayat[1], Cahyadi[2], Hafidz Sanjaya[3], Dwi Purnomo[4], Heri Wiranto[5]
[1,2,3,4,5]*Universitas Yayasan Pendidikan Imam Bonjol Majalengka, Indonesia*
Responden: fajarmaulahidayat@univypib.ac.id

ABSTRAK
Penelitian ini membahas penerapan metode *Latent Dirichlet Allocation* (LDA) untuk pemodelan topik berita terkini di Indonesia. Data dikumpulkan melalui RSS feed dari beberapa portal berita nasional seperti Detik, Kompas, Tribunnews, Liputan6, Tempo, CNN Indonesia, dan Antara News. Proses penelitian meliputi tahapan pengambilan data, pembersihan dan *preprocessing* teks, eksplorasi awal frekuensi kata, penyusunan representasi korpus, pemodelan topik LDA, visualisasi interaktif dengan pyLDAvis, serta evaluasi model menggunakan metrik *coherence score*. Hasil analisis menunjukkan model LDA dengan lima topik memberikan distribusi kata kunci yang relevan dengan isu-isu utama seperti bencana, politik, demonstrasi, korupsi, dan kriminal. Nilai *coherence score* sebesar 0,3591 mengindikasikan tingkat koherensi cukup baik, meskipun terdapat ruang optimasi melalui penyesuaian parameter. Visualisasi interaktif menunjukkan keterpisahan topik yang memadai, dengan tumpang tindih yang relatif kecil. Temuan ini memperlihatkan bahwa LDA efektif untuk mengidentifikasi topik dominan dalam berita nasional, sehingga dapat dimanfaatkan untuk analisis tren isu publik, pengelompokan konten media, serta mendukung pengambilan keputusan berbasis data.

***Kata Kunci**: topic modeling, LDA, berita Indonesia, coherence score, NLP.*

Riwayat Artikel :

Tanggal diterima : 13-12-2025 Tanggal terbit : 22-01-2026

Kutipan :

Hidayat, F. M., Cahyadi, Sanjaya, H., Purnomo, D., & Wiranto, H. (2026). PEMODELAN TOPIK BERITA NASIONAL INDONESIA MENGGUNAKAN LATENT DIRICHLET ALLOCATION (LDA). INFOTECH Journal, 12(1), 33–39. https://doi.org/10.31949/infotech.v12i1.16926

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong peningkatan signifikan dalam konsumsi berita digital, termasuk di Indonesia, yang berimplikasi pada melimpahnya data teks dari berbagai portal daring (Bamasputri et al., 2025). Kondisi ini menuntut adanya metode analisis yang efisien untuk mengungkap pola serta informasi tersembunyi, sehingga pemahaman terhadap topik utama dalam pemberitaan menjadi penting bagi pemerintah, akademisi, pelaku industri, dan masyarakat dalam mengidentifikasi isu-isu aktual, tren sosial, serta dinamika opini publik (Suardi, 2025). Realitas yang diberitakan oleh media massa kerap menjadi rujukan masyarakat dalam memahami peristiwa (Zahra et al., 2020), sehingga media memiliki tanggung jawab untuk menjaga kebebasan penyampaian informasi, menyajikan keragaman konten, serta memastikan kualitas informasi yang disampaikan dengan tetap menghormati hak-hak individu dan hak asasi manusia (Bakhtiar & Bima, 2020).

Seiring dengan terus meningkatnya volume berita, analisis manual menjadi tidak lagi efektif sehingga diperlukan pendekatan komputasi seperti *topic modeling* untuk mengekstraksi tema dominan secara otomatis dari kumpulan teks yang besar (Rakhmawati et al., 2024). *Topic modeling* merupakan salah satu teknik *unsupervised machine learning* yang berfungsi mengidentifikasi tema tersembunyi dari korpus dokumen berukuran besar dan mengelompokkan tema-tema tersebut ke dalam topik tertentu (Hosseiny Marani & Baumer, 2023; Kherwa & Bansal, 2020). Salah satu metode yang paling popular adalah *Latent Dirichlet Allocation* (LDA) yang pertama kali diperkenalkan oleh David Blei. LDA menggunakan pendekatan probabilistik pada distribusi kata dalam dokumen untuk merepresentasikan topik, sehingga dapat membantu menyajikan dokumen menjadi lebih terstruktur (Farkhod et al., 2021). Penerapan *topic modelling* dengan LDA dinilai unggul karena mampu menghasilkan topik yang bermakna secara logis (Albalawi et al., 2020), memiliki kemampuan interpretasi yang baik, serta performa prediktif yang lebih optimal (Dieng et al., 2020).

Berbagai penelitian sebelumnya telah mengimplementasi *topic modelling* dengan beragam konteks. H. T. Saputra et al. (2025), misalnya memperkenalkan *Latent Dirichlet Allocation* (LDA) melalui penelitian berjudul Analisis Modelling pada Reviews Lazada Indonesia Menggunakan *Latent Dirichlet Allocation* (LDA) untuk Optimalisasi Strategi Bisnis. Penelitian lain oleh Suhaeni et al. (2025) membahas *LDA Topic Modeling Analysis of Public Discourse on Indonesia's Free Nutritious Meals Program* (MBG). Selanjutnya, Ramadhani et al. (2025) menerapkan LDA pada Evaluasi Aplikasi Layanan Alfagift Berdasarkan Topik Ulasan Pengguna, sementara Wahyuni et al. (2025) memanfaatkan *BERTopic* melalui kajian *Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments*. Meskipun demikian, sebagian besar penelitian tersebut cenderung berfokus pada isu-isu spesifik seperti program makan bergizi gratis, *e-commerce*, atau aplikasi tertentu. Belum banyak kajian yang secara komprehensif menyoroti topik dominan dalam berita daring di Indonesia pada periode terkini dengan pendekatan sistematis.

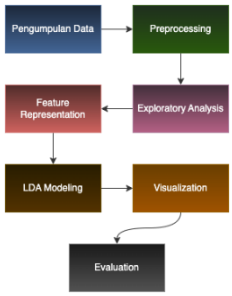
Berdasarkan kajian tersebut, terdapat beberapa celah penelitian. Pertama, sebagian besar studi membatasi cakupan pada dominan isu tertentu sehingga belum memberikan gambaran menyeluruh mengenai spektrum topik yang diberitakan media. Kedua, beberapa penelitian menggunakan dataset lama atau rentang waktu yang terbatas, sehingga kurang merepresentasikan dinamika isu terkini yang sangat cepat berubah. Ketiga, evaluasi kualitas topik, misalnya melalui *coherence score*, jarang dilakukan secara eksplisit, padahal metrik ini penting untuk menilai keandalan hasil pemodelan topik.

Dengan demikian, diperlukan penelitian yang mampu mengekstraksi dan menganalisis topik dominan pada berita daring Indonesia dari berbagai sumber kredibel dalam periode waktu terbaru. Penelitian ini diharapkan dapat memberikan kontribusi dalam memahami tren isu masyarakat sekaligus menjadi dasar pengambilan kebijakan berbasis data. Adapun tujuan penelitian ini adalah menganalisis topik-topik dominan yang muncul dalam berita daring menggunakan metode LDA, mengevaluasi kualitas model topik melalui perhitungan *coherence score*, serta mengidentifikasi topik yang paling sering muncul sebagai dasar untuk analisis lanjutan.

2. METODE

a. Desain Penelitian

Pendekatan ini menggunakan pendekatan eksperimental dengan tujuan mengidentifikasi topik-topik dominan dalam berita daring melalui metode *Latent Dirichlet Allocation* (LDA) (Rakhmawati et al., 2024). Seluruh tahapan analisis dilakukan menggunakan bahasa pemrograman Python pada lingkungan Google Colab, sehingga seluruh proses dapat direplikasi secara mudah dan transparan. Untuk memperjelas tahapan analisis, Gambar 1 menampilkan alur penelitian yang digunakan dalam penelitian ini.



Gambar 1. Model penelitian

b. Pengumpulan Data

Dataset yang dianalisis terdiri atas kumpulan berita daring selama tiga bulan terakhir yang didapat dari beberapa portal berita nasional, seperti Detik, Kompas, Tribunnews, Liputan6, Tempo, CNN Indonesia, dan Antara News. Data dikumpulkan dan disimpan dalam format CSV dengan variabel utama berupa judul serta deskripsi/isi berita. Selanjutnya, dataset diunggah ke Google Colab menggunakan pandas.read_csv() untuk keperluan pemrosesan lebih lanjut.

c. Prosedur Penelitian

Proses pengolahan data dan pemodelan topik dilakukan melalui beberapa tahapan:

1) Persiapan Lingkungan dan Pustaka

Instalasi dan pemanggilan pustaka pendukung meliputi *pandas, numpy, matplotlib, seaborn, nltk, gensim, pyLDAvis,* dan *wordcloud*. Selain itu, daftar *stopwords* untuk bahasa Indonesia dan Inggris diunduh melalui *nltk* untuk mendukung tahap *preprocessing* teks.

2) *Preprocessing* Teks

Pada tahap ini, variabel judul dan deskripsi berita terlebih dahulu digabungkan menjadi satu teks utuh agar informasi yang dianalisis lebih komprehensif. Teks kemudian dinormalisasi dengan mengubah seluruh huruf menjadi bentuk *lowercase* (*case folding*) serta menghapus karakter non-alfabet menggunakan *regular expression* sebagai bagian dari proses pembersihan teks (*cleaning text*) (Khairani et al., 2024; Nugroho et al., 2025). Selanjutnya, dilakukan tokenisasi untuk memecah teks menjadi unit kata (*tokenization*) (M. Saputra & Sri Wahyuni, 2024; Sari et al., 2023), diikuti dengan penghapusan *stopwords* baik dalam bahasa Indonesia dan Inggris (*stopwords removal*) (Slamet et al., 2022), serta eliminasi kata yang memiliki panjang kurang dari dua huruf. Hasil akhir dari tahapan ini adalah kumpulan token yang sudah bersih dan siap digunakan dalam analisis lebih lanjut.

3) Eksplorasi Awal dan Visualisasi Frekuensi Kata

Distribusi kata dihitung menggunakan pustaka *Counter* untuk mengidentifikasi kata-kata dengan frekuensi tertinggi. Dua puluh kata dengan jumlah kemunculan terbesar divisualisasikan menggunakan *WordCloud* guna memberikan gambaran intuitif mengenai kata-kata yang paling menonjol (Efendi et al., 2025).

4) Penyusunan Representasi untuk LDA

Representasi data untuk LDA disusun dengan membangun *dictionary* dan *corpus* menggunakan gensim.corpora.Dictionary dan fungsi doc2bow. *Dictionary* menyimpan daftar kata unik, sedangkan *corpus* berisi kata beserta jumlah kemunculannya di setiap dokumen. Pada tahap ini juga ditampilkan jumlah dokumen serta kata unik untuk memberikan gambaran dimensi data.

5) Pemodelan Topik LDA

Tahap inti penelitian adalah pemodelan topik menggunakan LDA. Jumlah topik yang diekstraksi ditentukan, misalnya lima topik, kemudian model LDA dilatih dengan gensim.models.LdaModel. Parameter pelatihan mencakup passes=10 (jumlah iterasi) dan random_state=42 untuk memastikan hasil yang dapat direplikasi. *Output* berupa daftar topik beserta kata-kata penyusunnya yang merepresentasikan karakteristik masing-masing topik (Yu & Xiang, 2023).

6) Visualisasi Interaktif Topik

Hasil pemodelan LDA divisualisasikan menggunakan pyLDAvis untuk melihat distribusi dan kata kunci dari setiap topik secara interaktif. Visualisasi ini membantu menilai keterpisahan serta dominasi topik dalam korpus dan memudahkan interpretasi hasil.

d. Metrik Evaluasi

Kualitas topik yang dihasilkan dievaluasi dengan menggunakan metrik koherensi (*coherence score*) berbasis *c_v* yang dihitung melalui gensim.models.CoherenceModel. Metrik ini mengukur sejauh mana kata-kata penyusun topik melalui keterkaitan makna yang konsisten, sehingga memudahkan interpretasi pada tahap analisis (Marzuqi et al., 2025). Selain itu, distribusi topik divisualisasikan untuk menilai sebaran dan keterpisahan antar-topik dalam korpus. Analisis ini memastikan bahwa setiap topik memiliki diferensiasi yang jelas dan tidak tumpang tindih, sehingga hasil pemodelan dapat dinilai secara komprehensif.

3. PEMBAHASAN

a. Penyajian Hasil dan Analisis

1) Persiapan Lingkungan dan Pustaka

Proses analisis dilakukan di lingkungan Google Colab yang mendukung komputasi berbasis *cloud*, sehingga tidak memerlukan sumber daya lokal yang besar. Pada tahap awal, seluruh pustaka yang dibutuhkan diinstall dan diimpor, seperti pandas untuk pengolahan data tabular, gensim untuk pemodelan topik LDA, serta pyLDAvis untuk visualisasi interaktif topik. Pemilihan Google Colab juga memberikan kemudahan replikasi karena skrip data

dijalankan langsung di *cloud* tanpa konfigurasi yang rumit.

	tanggal	judul	deskripsi	Link	sumber
0	2025-08-25	Polisi Ungkap Situasi di Perbatasan RI-RDTL se...	<img src="https://akcdn.detik.net.id/visual/20...	https://www.cnnindonesia.com/nasional/20250825...	https://www.cnnindonesia.com/nasional/rss
1	2025-08-25	Penjelasan Pramono soal Alih Fungsi Trotoar di...	<img src="https://akcdn.detik.net.id/visual/20...	https://www.cnnindonesia.com/nasional/20250825...	https://www.cnnindonesia.com/nasional/rss
2	2025-08-25	Kepala BNN Beber Kans Indonesia Larang Vape se...	<img src="https://akcdn.detik.net.id/visual/20...	https://www.cnnindonesia.com/nasional/20250825...	https://www.cnnindonesia.com/nasional/rss
3	2025-08-25	Sempat Ditutup, Pejompongan dan Tol Dalkot Sem...	<img src="https://akcdn.detik.net.id/visual/20...	https://www.cnnindonesia.com/nasional/20250825...	https://www.cnnindonesia.com/nasional/rss
4	2025-08-25	Bupati Sebut Warga NTT Protes Patok Kena Temba...	<img src="https://akcdn.detik.net.id/visual/20...	https://www.cnnindonesia.com/nasional/20250825...	https://www.cnnindonesia.com/nasional/rss

Gambar 2. Dataset

Gambar 2 menunjukkan awal dataset yang telah diunggah ke Colab dalam format CSV. Dataset terdiri dari beberapa kolom atau variabel seperti tanggal, judul berita, deskripsi/isi berita, dan tautan sumber. Dataset tersebut mengilustrasikan struktur data mentah sebelum dilakukan tahap *preprocessing*. Keberadaan metadata seperti tanggal dan sumber berita penting untuk analisis lanjutan karena dapat membantu mengaitkan topik dengan konteks waktu serta sumber publikasi. Selain itu, dataset diperoleh secara langsung melalui RSS feed dari portal berita nasional terpercaya seperti Detik, Kompas, Tribunnews, Liputan6, Tempo, CNN Indonesia, dan Antara News, sehingga validitas dan keaslian data dapat dipertanggungjawabkan.

2) Preprocessing Teks

Tahap *preprocessing* dilakukan untuk memastikan bahwa data teks yang dianalisis berada dalam kondisi bersih dan siap di proses oleh model LDA. Proses ini meliputi sejumlah langkah penting, dimulai normalisasi teks melalui pengubahan seluruh huruf menjadi format *lowercase*, penghapusan karakter non-alfabet seperti tanda baca dan angka, serta pembersihan elemen yang tidak relevan seperti tag HTML, URL, dan jalur file gambar. Selanjutnya dilakukan tokenisasi, yaitu pemecahan teks berita menjadi unit kata (*tokens*) berdasarkan spasi.

Proses *preprocessing* juga mencakup penghapusan *stopwords* baik dalam Bahasa Indonesia maupun bahasa Inggris serta kata-kata yang memiliki panjang ≤ 2 huruf. Langkah ini bertujuan untuk mengurangi *noise* sehingga hanya kata-kata bermakna yang dipertahankan untuk analisis lebih lanjut. Hasil dari tahap ini adalah daftar token yang lebih representatif dan dapat menjadi dasar yang baik untuk pembentukan topik.

	text	tokens
0	Polisi Ungkap Situasi di Perbatasan RI-RDTL se...	[polisi, situasi, perbatasan, rirdtl, penembak...
1	Penjelasan Pramono soal Alih Fungsi Trotoar di...	[penjelasan, pramono, alih, fungsi, trotoar, j...
2	Kepala BNN Beber Kans Indonesia Larang Vape se...	[kepala, bnn, beber, kans, indonesia, larang, ...
3	Sempat Ditutup, Pejompongan dan Tol Dalkot Sem...	[ditutup, pejompongan, tol, dalkot, semanggisi...
4	Bupati Sebut Warga NTT Protes Patok Kena Temba...	[bupati, warga, ntt, protes, patok, kena, temb...

Gambar 3. Dataset setelah proses preprocessing

Contoh hasil tokenisasi awal dapat dilihat pada Gambar 3, yang memperlihatkan konversi dari teks berita mentah menjadi daftar token bersih. Misalnya, teks berita dengan judul dan isi tertentu dipetakan menjadi kata-kata kunci seperti polisi, situasi, perbatasan, penembak dan seterusnya.

Interpretasi dari tahap ini menegaskan bahwa kualitas token memiliki peran penting terhadap keberhasilan model LDA. Apabila pembersihan dan tokenisasi tidak dilakukan dengan cermat, model dapat menghasilkan topik yang bias, ambigu, atau sulit diinterpretasikan.

3) Eksplorasi Awal dan Frekuensi Kata

Tahap eksplorasi awal dilakukan untuk melihat distribusi kata dalam korpus berita setelah proses *preprocessing*. Analisis dilakukan dengan menghitung frekuensi kemunculan kata dan divisualisasikan menggunakan *WordCloud* serta daftar kata dengan frekuensi tertinggi.

Gambar 4. Frekuensi kata yang sering muncul

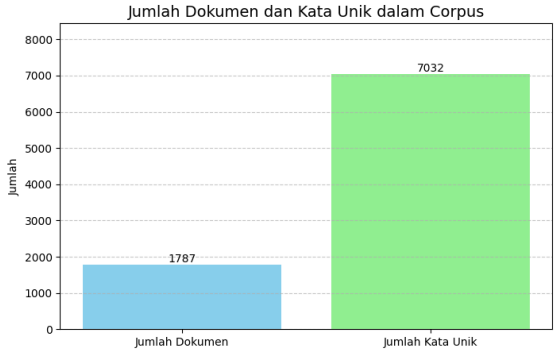
Hasil visualisasi pada Gambar 4 menunjukkan kata-kata dominan seperti “indonesia”, “sdn kalibaru”, “polisi”, “kpk”, “dpr”, “sea games”, dan “pemerintah”, yang mengindikasikan fokus pemberitaan pada isu hukum, politik, dan pemerintahan selama periode pengambilan data. Fenomena ini sejalan dengan dinamika politik dan kasus-kasus hukum yang banyak mendapat sorotan publik, serta konsisten dengan temuan penelitian terdahulu yang menyebutkan bahwa media daring Indonesia cenderung menyoroti isu serupa pada masa ketidakstabilan sosial.

Visualisasi distribusi kata ini penting sebagai dasar awal untuk memahami tema dominan pada korpus sebelum dilakukan pemodelan topik LDA, sehingga hasil topik yang terbentuk dapat lebih mudah diinterpretasikan.

4) Penyusunan Representasi untuk LDA

Tahap berikutnya adalah penyusunan representasi data untuk pemodelan topik LDA. Proses ini dilakukan dengan

membentuk *dictionary* dan *corpus* dari hasil *preprocessing*. Seperti ditunjukkan pada Gambar 5, jumlah dokumen yang diolah sebanyak 1.787, sedangkan jumlah kata unik yang terbentuk mencapai 7.032.



Gambar 5. Jumlah dokumen dan kata unik

Tingginya jumlah kata unik menunjukkan keragaman kosakata yang cukup baik, namun tetap terkendali berkat pembersihan pada tahap *preprocessing*. Keragaman ini penting karena menjadi landasan pembentukan topik yang lebih bervariasi, tetapi tetap relevan dengan konteks berita yang dianalisis. Dengan representasi ini, model LDA dapat mengidentifikasi pola dan keterkaitan kata yang membentuk topik secara lebih optimal.

5) Pemodelan Topik LDA

Model LDA diterapkan dengan jumlah 5 topik, menghasilkan daftar kata kunci dominan untuk setiap topik seperti yang ditunjukkan Gambar 6. Setiap topik merepresentasikan isu utama yang muncul pada korpus berita.

```
Topik 0: 0.011*"indonesia" + 0.010*"banjir" + 0.009*"games"
Topik 1: 0.020*"prabowo" + 0.011*"presiden" + 0.006*"korups
Topik 2: 0.026*"dpr" + 0.017*"demo" + 0.010*"agustus" + 0.0
Topik 3: 0.015*"kpk" + 0.013*"haji" + 0.010*"bupati" + 0.00
Topik 4: 0.019*"mobil" + 0.014*"siswa" + 0.013*"mbg" + 0.01
```

Gambar 6. Proses penentuan topik menggunakan LDA

Sebagai contoh, Topik 1 berhubungan dengan isu politik dengan kata dominan seperti prabowo, presiden, korupsi, bandung, dan subianto. Topik 2 berkaitan dengan isu korupsi seperti kpk, haji, bupati, lampung, dan menteri. Topik lainnya mencakup berbagai aspek seperti kriminal, bencana, dan demonstrasi.

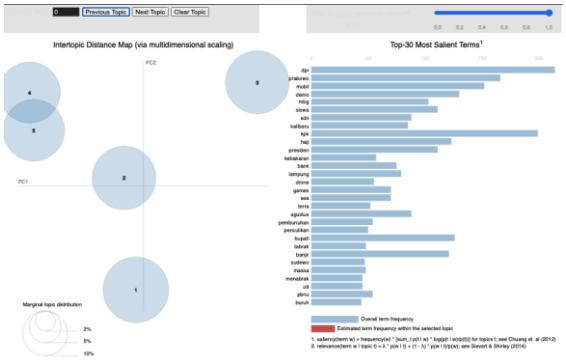
Hasil ini menunjukkan bahwa fokus pemberitaan selama periode pengambilan data cenderung tertuju pada isu politik dan korupsi, sementara isu lain seperti kriminal, bencana, dan demonstrasi tetap mendapat perhatian meskipun tidak sekuat dua topik utama tersebut.

6) Visualisasi Interaktif

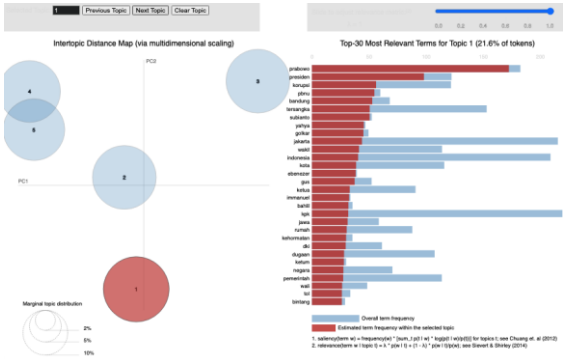
Visualisasi interaktif menggunakan pyLDAvis seperti yang ditunjukkan pada Gambar 7 sampai Gambar 12, yang memberikan gambaran distribusi dan keterpisahan antar topik. Lingkaran pada panel kiri merepresentasikan setiap topik, sementara panel kanan menampilkan kata kunci dengan kontribusi terbesar pada masing-masing topik.

Secara umum, lingkaran topik terlihat relative terpisah, yang menunjukkan bahwal model LDA berhasil membentuk topik dengan keterbedaan makna yang cukup baik. Pada beberapa area terdapat sedikit tumpang tindih, namun hal ini wajar mengingat adanya keterkaitan kontekstual antar isu.

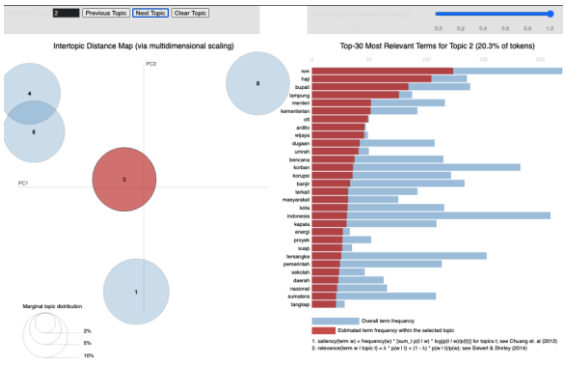
Visualisasi ini membantu peneliti mengevaluasi kualitas model LDA secara intuitif. Dengan keterpisahan topik yang baik, interpretasi terhadap tema dominan seperti isu politik (topik 1), korupsi (topik 2), kriminal (topik 3), bencana (topik 4), dan demonstrasi (topik 5) menjadi lebih jelas dan mendukung temuan pada tahap sebelumnya.



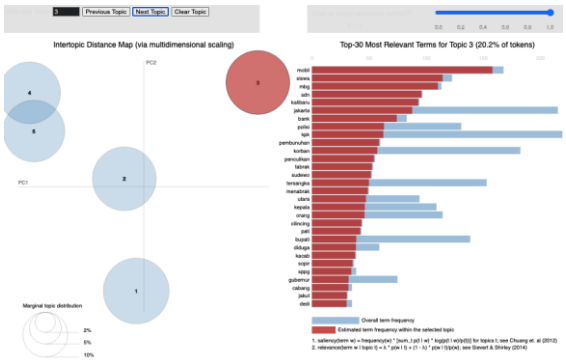
Gambar 7. Semua topik



Gambar 8. Topik 1 isu Politik

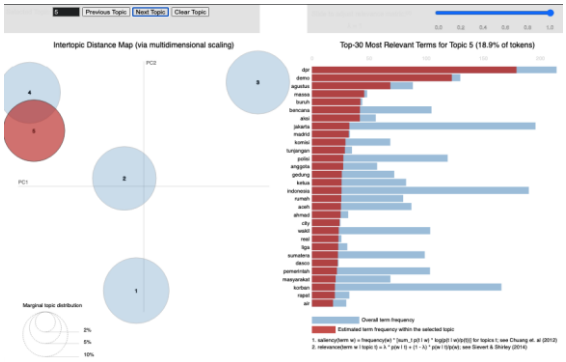


Gambar 9. Topik 2 isu Korupsi



Gambar 10. Topik 3 isu Kriminal

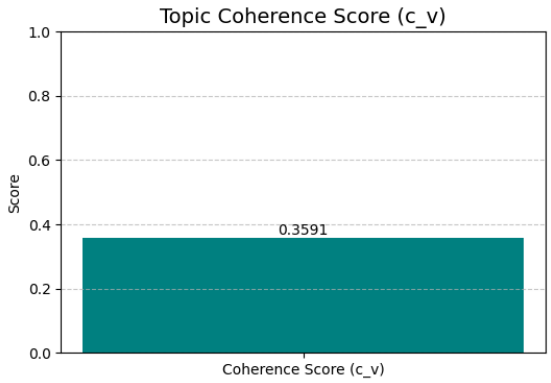
Gambar 11. Topik 4 isu Bencana



Gambar 12. Topik 5 isu Demonstrasi

b. Metrik Evaluasi

Model LDA dengan lima topik menghasilkan *coherence score* sebesar 0,3591 seperti yang ditunjukkan pada Gambar 13. Nilai ini menunjukkan tingkat koherensi yang cukup baik, di mana kata-kata penyusun setiap topik memiliki keterkaitan semantik yang relevan sehingga topik yang dihasilkan masih dapat diinterpretasikan dengan jelas. Meskipun demikian, nilai tersebut mengindikasikan bahwa masih terdapat ruang untuk optimasi, misalnya melalui penyesuaian jumlah topik, penambahan jumlah iterasi (*passes*), atau penyempurnaan tahap *preprocessing* guna meningkatkan kualitas koherensi model.

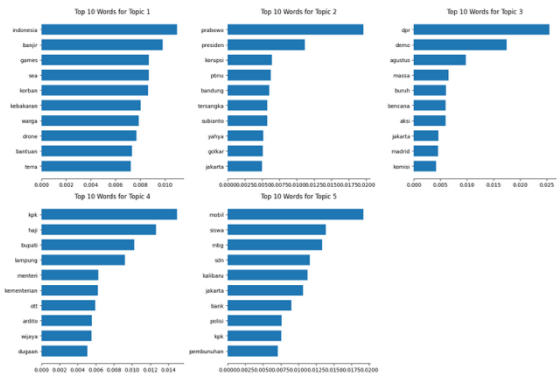


Gambar 13. Coherence score

Jika dibandingkan dengan penelitian sejenis, nilai *coherence score* pada penelitian ini berada pada kategori sedang. Penelitian oleh Cici Suhaeni, dkk (2025) yang berfokus pada analisis topik program Makan Bergizi Gratis (MBG) memperoleh nilai koherensi sebesar 0,5721, sementara penelitian Haris Tri Saputra, dkk (2025) pada ulasan Lazada Indonesia memperoleh nilai koherensi yang lebih tinggi, yaitu 0,7179. Perbedaan ini dapat dipengaruhi oleh karakteristik dataset yang digunakan, di mana kedua penelitian tersebut memiliki fokus topik yang lebih spesifik dan homogen dibandingkan dengan penelitian ini yang menganalisis berita nasional Indonesia secara umum dan mencakup berbagai isu lintas domain.

Keragaman topik dalam korpus berita nasional menyebabkan distribusi kata yang lebih kompleks dan beragam, sehingga berdampak pada nilai koherensi yang relatif lebih rendah. Namun demikian, kondisi ini mencerminkan realitas pemberitaan nasional yang dinamis dan multidimensi. Dengan demikian, nilai *coherence score* sebesar 0,3591 masih dapat dianggap wajar dan representatif untuk dataset berita yang bersifat umum dan heterogen.

Analisis kata kunci pada Gambar 14 menunjukkan bahwa setiap topik memiliki identitas yang jelas, seperti isu bencana, politik, demonstrasi, korupsi, dan kriminal. Distribusi topik dalam korpus relatif seimbang, meskipun terdapat beberapa dokumen yang lebih dominan pada topik tertentu. Jarak antar topik dalam visualisasi terlihat cukup terpisah, yang mengindikasikan bahwa model LDA mampu membedakan tema utama dengan baik dan mendukung proses interpretasi hasil.



Gambar 14. Hasil topic modeling menggunakan LDA

Secara keseluruhan, kombinasi antara nilai *coherence score* dan visualisasi distribusi topik menunjukkan bahwa model LDA yang dibangun telah efektif dalam mengelompokkan isu-isu utama pada berita nasional Indonesia. Ke depan, evaluasi lanjutan dengan variasi parameter maupun penggunaan metrik tambahan dapat dilakukan untuk memperoleh kualitas model yang lebih optimal.

4. KESIMPULAN

Penelitian ini berhasil mengidentifikasi topik utama dari berita Indonesia terkini menggunakan metode *Latent Dirichlet Allocation* (LDA), yang menunjukkan bahwa berita nasional didominasi oleh isu bencana, politik, demonstrasi, korupsi, dan kriminal. Hasil pemodelan topik ini memberikan gambaran yang komprehensif mengenai kecenderungan informasi yang beredar pada portal berita nasional, meskipun karakteristik data yang bersifat heterogen mempengaruhi tingkat koherensi topik yang dihasilkan.

Berdasarkan evaluasi menggunakan *coherence score*, model LDA menunjukkan kualitas pemodelan yang cukup baik dan mampu membedakan tema utama secara jelas. Temuan ini mendukung tujuan penelitian dalam menyajikan analisis topik yang terstruktur dan objektif. Dengan demikian, hasil penelitian ini dapat menjadi dasar bagi pengembangan kajian lanjutan, seperti analisis sentimen, pemantauan isu publik, atau eksplorasi konten media berbasis metode pembelajaran mesin.

PUSTAKA

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, *3*, 42.

Bakhtiar, N. Y., & Bima, M. R. (2020). Pemenuhan Hak Kebebasan Berpendapat Berdasarkan Undang-undang Nomor 9 Tahun 1999 Tentang Kemerdekaan Berpendapat Di Muka Umum. *Journal of Lex Theory (JLT)*, *1*(1), 41–58.

Bamasputri, R. A., Fitriani, I., Salsabila, A. A., & Prasetyaningsih, N. I. D. (2025). TRANSFORMASI DIGITAL: EVOLUSI KOMUNIKASI MASSA DI ERA MEDIA SOSIAL DAN DAMPAKNYA PADA PERILAKU KONSUMSI INFORMASI MASYARAKAT. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, *11*(02), 270–286.

Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439–453.

Efendi, N. M., Muhammad Ryan Adam Saputra, Srikandi, D., Girsang, S., Rizky Maulana Dzuhry, & Mohamad Andi Budiono. (2025). IMPLEMENTASI ALGORITMA NAÏVE BAYES PADA SENTIMEN PUBLIK TERHADAP SOLUSI MENGHADAPI RESESI DI INDONESIA. *INFOTECH Journal*, *11*(2), 451–457. https://doi.org/10.31949/infotech.v11i2.16723

Farkhod, A., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model. *Applied Sciences*, *11*(23), 11091.

Hosseiny Marani, A., & Baumer, E. P. S. (2023). A review of stability in topic modeling: Metrics for assessing and techniques for improving stability. *ACM Computing Surveys*, *56*(5), 1–32.

Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(4), 887–894. https://doi.org/10.25126/jtiik.1148315

Kherwa, P., & Bansal, P. (2020). Topic modeling: a comprehensive review. *EAI Endorsed Trans. Scalable Inf. Syst.*, *7*(24), e2.

Marzuqi, F., Purba, Y., Syaputra, T., Bismi, W., Kurniawati, I., & Fahlapi, R. (2025). PERBANDINGAN ALGORITMA MACHINE LEARNING UNTUK SENTIMEN ULASAN APLIKASI JOBSTREET. *INFOTECH Journal*, *11*(2), 416–423. https://doi.org/10.31949/infotech.v11i2.16609

Nugroho, D. C., Agustina, A. D., Maulana, B., Darussalam, F., Asyraf, M., & Lubis, B. O. (2025). PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PUBLIK PADA PROGRAM MAKAN BERGIZI GRATIS. *INFOTECH Journal*, *11*(2), 474–479. https://doi.org/10.31949/infotech.v11i2.16781

Rakhmawati, N. A., Cisatra, A., Ansori, D. D. M., Akmal, D. N. F. A., & Ramadhani, S. (2024). Identifikasi Topik Hangat di Media Berita Menggunakan Latent Dirichlet Allocation. *JIEET (Journal of Information Engineering and Educational Technology)*, *8*(1), 14–17.

Ramadhani, T. Z., Setiawan, N. Y., & Rahayudi, B. (2025). Evaluasi Aplikasi Layanan Alfagift Berdasarkan Topik Ulasan Pengguna

Menggunakan Latent Dirichled Allocation (LDA). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *9*(6).

Saputra, H. T., Damanik, A., Shaquille, M. M., Rusydi, M. A., Riziq, M. H., & Zulva, N. S. (2025). Analisis Modelling Pada Reviewes Lazada Indonesia Menggunakan Latent Dirichlet Allocation (LDA) Untuk Optimalisasi Strategi Bisnis. *JEKIN-Jurnal Teknik Informatika*, *5*(1), 361–371.

Saputra, M., & Sri Wahyuni. (2024). ANALISIS SENTIMEN PENGGUNA PADA APLIKASI BANK DIGITAL KROM DENGAN ALGORITMA SUPPORT VECTOR MACHINE. *INFOTECH Journal*, *10*(2), 327–332. https://doi.org/10.31949/infotech.v10i2.11801

Sari, T. A., Sinduningrum, E., & Hasan, F. N. (2023). Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naïve Bayes. *KLIK Kaji. Ilm. Inform. Dan Komput*, *3*(6), 773–779.

Slamet, R., Gata, W., Novtariany, A., Hilyati, K., & Jariyah, F. A. (2022). Analisis sentimen Twitter terhadap penggunaan artis Korea Selatan sebagai brand ambassador produk kecantikan lokal. *INTECOMS J. Inf. Technol. Comput. Sci, 5*(1), 145–153.

Suardi, S. (2025). MENINGKATKAN KREDIBILITAS MEDIA DI INDONESIA DALAM ERA DISRUPSI INFORMASI: STRATEGI MENGHADAPI MISINFORMASI DIGITAL. *Jurnal Ilmu Komunikasi UHO: Jurnal Penelitian Kajian Ilmu Komunikasi Dan Informasi*, *10*(1), 249–258.

Suhaeni, C., Mualifah, L. N. A., & 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.

Wahyuni, W., Lestari, T. P., Apriliana, M., & Gumelta, R. (2025). Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments. *Journal of Applied Informatics and Computing*, *9*(4), 1964–1971.

Yu, D., & Xiang, B. (2023). Discovering topics and trends in the field of Artificial Intelligence: Using LDA topic modeling. *Expert Systems with Applications*, *225*, 120114. https://doi.org/10.1016/j.eswa.2023.120114

Zahra, F., Mustaqimmah, N., & Hendra, M. D. (2020). Kekuatan Media Digital Pada Pembentukan Budaya Populer (Studi Pada Komunitas Moarmy Pekanbaru). *Komunikasiana: Journal of Communication Studies*, *2*(2), 109–122.