

---

---

## KLASTERISASI TOPIK PEMBERITAAN BEA CUKAI PADA MEDIA SOSIAL MENGGUNAKAN LATENT DIRICHLET ALLOCATION (LDA) PASCA PERGANTIAN MENTERI KEUANGAN 2025

Irwan<sup>1</sup>, Dwina Pri Indini<sup>2</sup>, Feby Wulandari Sembiring<sup>3</sup>, Vina Gusti Malem<sup>4</sup>,  
Daslan Fernando Aritonang<sup>5</sup>

Universitas Pembangunan Panca Budi, Medan

e-mail: irwanbc79@gmail.com<sup>1</sup>, dwinapriindini03@gmail.com<sup>2</sup>,

febywulandari421@gmail.com<sup>3</sup>, vinahach6@gmail.com<sup>4</sup>, daslan13@gmail.com<sup>5</sup>

**Abstract:** *The replacement of Indonesia's Minister of Finance from Sri Mulyani Indrawati to Purbaya Yudhi Sadewa on September 8, 2025, triggered significant public discourse regarding institutions under the Ministry of Finance, particularly the Directorate General of Customs and Excise (DJBC). This research aims to identify and classify dominant topics discussed by the public on social media regarding Customs during the period from September 2025 to April 2026 using the Latent Dirichlet Allocation (LDA) algorithm. A total of 937 public comments were collected through web scraping on YouTube from 15 relevant news videos, followed by preprocessing including case folding, tokenization, stopword removal using the Sastrawi library, and stemming. The optimal number of topics was determined by evaluating Coherence Score (C<sub>v</sub>) and Log Perplexity across the range  $K = 2$  to  $K = 15$ . The results indicate that the highest coherence value was obtained at  $K = 11$  with a score of 0.3436, which was selected as the final model. The eleven topics generated reflect a spectrum of public opinion covering appreciation for Minister Purbaya, demands for cleansing corrupt officials, illegal smuggling issues, and demands for institutional reform. Topic distribution shows that 52.3% of documents were critical, 29.9% supportive, and 18.1% reform-demanding. These findings can serve as a basis for evaluating DJBC's public communication policies under Minister Purbaya's leadership.*

**Keywords:** *latent dirichlet allocation; topic modeling; customs; social media; coherence score.*

**Abstrak:** Pergantian Menteri Keuangan Republik Indonesia dari Sri Mulyani Indrawati kepada Purbaya Yudhi Sadewa pada 8 September 2025 memicu dinamika pemberitaan signifikan terhadap institusi-institusi di bawah Kementerian Keuangan, khususnya Direktorat Jenderal Bea dan Cukai (DJBC). Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan topik-topik dominan yang diperbincangkan publik di media sosial terkait Bea Cukai pada periode September 2025 hingga April 2026 menggunakan algoritma Latent Dirichlet Allocation (LDA). Data sejumlah 937 komentar publik dikumpulkan melalui teknik web scraping pada platform YouTube dari 15 video pemberitaan relevan, kemudian diolah melalui tahapan preprocessing berupa case folding, tokenisasi, stopword removal menggunakan library Sastrawi, dan stemming. Penentuan jumlah topik optimal dilakukan dengan evaluasi Coherence Score (C<sub>v</sub>) dan Log Perplexity pada rentang  $K = 2$  hingga  $K = 15$ . Hasil penelitian menunjukkan bahwa nilai coherence tertinggi diperoleh pada  $K = 11$  dengan skor 0,3436, yang selanjutnya dijadikan sebagai model final. Sebelas topik yang dihasilkan mencerminkan spektrum opini publik yang terdiri dari topik apresiasi terhadap Menkeu Purbaya, tuntutan pembersihan oknum korup, isu penyelundupan barang ilegal, serta desakan reformasi kelembagaan. Distribusi topik menunjukkan bahwa 52,3% dokumen bersifat kritis, 29,9% bersifat dukungan, dan 18,1% berupa desakan reformasi. Temuan ini dapat dimanfaatkan sebagai dasar evaluasi kebijakan komunikasi publik DJBC di era kepemimpinan Menkeu Purbaya.

**Kata kunci:** bea cukai; coherence score; latent dirichlet allocation; media sosial; pemodelan topik

## PENDAHULUAN

Media sosial telah menjadi ruang publik digital yang merepresentasikan opini, kritik, dan harapan masyarakat terhadap lembaga pemerintahan. Platform seperti YouTube, Twitter/X, dan Instagram memungkinkan publik menyampaikan pandangan secara masif dan real-time, menghasilkan volume data tekstual yang sangat besar dan kaya makna. Dalam konteks kebijakan publik, data ini menyimpan informasi berharga yang dapat diolah untuk memahami persepsi masyarakat terhadap kinerja institusi negara (Rakhmawati et al., 2024).

Direktorat Jenderal Bea dan Cukai (DJBC) merupakan salah satu institusi di bawah Kementerian Keuangan yang menjadi sorotan publik karena fungsinya sebagai penjaga perbatasan ekonomi negara. Berbagai peristiwa pemberitaan terkait DJBC mulai dari viralnya kasus penahanan barang impor, dugaan penyalahgunaan wewenang oknum, hingga kinerja pengawasan cukai kerap memicu gelombang opini di media sosial. Dinamika ini semakin intens pasca pergantian

Menteri Keuangan dari Sri Mulyani Indrawati kepada Purbaya Yudhi Sadewa pada 8 September 2025, yang membawa ekspektasi baru terhadap reformasi institusi keuangan negara (Farta Wijaya et al., 2025).

Analisis manual terhadap ribuan komentar publik tidak efisien dan rentan terhadap bias subjektif. Oleh karena itu, diperlukan pendekatan otomatis berbasis Natural Language Processing (NLP) untuk mengekstrak topik-topik dominan secara objektif (Yahya et al., 2025). Dalam upaya ekstraksi informasi tersebut, metode *Latent Dirichlet Allocation* (LDA) telah terbukti efektif sebagai algoritma *topic modeling* probabilistik yang mampu mengidentifikasi struktur

laten dalam kumpulan dokumen tekstual yang besar (Salwa, 2025). *Latent Dirichlet Allocation* (LDA), diperkenalkan oleh Blei, Ng, dan Jordan pada tahun 2003, merupakan salah satu metode *topic modeling* probabilistik paling populer dan terbukti efektif dalam mengidentifikasi struktur laten dari kumpulan dokumen tekstual (Sepriadi et al., 2025)

Pada penelitian sebelumnya berjudul “Analisis Persepsi Publik Terhadap Reshuffle Menteri Keuangan 2025 Berdasarkan Topic Modeling LDA Pada Komentar Youtube” (Wakhid et al., 2026) penelitian tersebut menyatakan bahwa dengan menggunakan algoritma LDA berhasil membentuk kelompok topik utama pada ribuan data komentar youtube dengan akurasi yang cukup baik sebesar 0,5 dan 0,8 pada *topic diversity*nya hal ini menegaskan bahwa model LDA berhasil mengelompokkan kualitas topik yang baik.

Penelitian selanjutnya berjudul “Pemodelan Topik dengan LDA Untuk Memahami Persepsi Publik Terhadap *Apple Vision Pro* Melalui Analisis Berita Twitter” mengungkapkan bahwa algoritma LDA berhasil mengidentifikasi topik utama dan mengelompokkan kedalam beberapa tema dengan interpretasi yang jelas, yang artinya algoritma LDA berhasil dan membantu memberikan wawasan yang bermakna tentang persepsi publik (Farhan et al., 2025).

Beberapa penelitian terdahulu telah mengaplikasikan LDA pada data media sosial di Indonesia, seperti analisis persepsi publik atau berita daring umum dan hanya menggunakan *coherence score* untuk menilai evaluasinya, namun berbeda dengan penelitian ini yang secara spesifik membedah transisi kepemimpinan di Kementerian Keuangan terhadap citra Bea Cukai masih sangat terbatas. Kondisi ini menciptakan sebuah

celah penelitian (*research gap*) mengenai bagaimana struktur opini publik terhadap citra bea cukai apakah berubah atau terbentuk tepat setelah pergantian pimpinan tertinggi kementerian tersebut.

Berdasarkan *research gap* dari penelitian terdahulu penelitian ini hadir dengan **kebaruan** (*novelty*) berupa analisis tematik terhadap 937 komentar YouTube yang dikumpulkan selama periode transisi kritis kepemimpinan Menkeu Purbaya. Berbeda dengan analisis sentimen sederhana, penelitian ini mengintegrasikan evaluasi teknis menggunakan *Coherence Score* dan *Log Perplexity* untuk memastikan bahwa topik yang dihasilkan tidak hanya akurat secara statistik tetapi juga memiliki interpretabilitas yang tinggi bagi pengambil kebijakan (Septiara et al., 2026).

Rumusan masalah penelitian ini adalah: (1) Bagaimana struktur topik dominan pemberitaan Bea Cukai di media sosial pasca pergantian Menkeu 2025? (2) Berapa jumlah topik optimal yang merepresentasikan data berdasarkan evaluasi *Coherence Score* dan *Perplexity*? (3) Bagaimana distribusi sentimen topik publik terhadap kinerja Bea Cukai di era Menkeu Purbaya?

Tujuan utama dari penelitian ini adalah untuk mengidentifikasi struktur topik dominan yang berkembang di media sosial terkait Bea Cukai dan menentukan jumlah topik paling optimal yang merepresentasikan keresahan publik. Dengan membedah distribusi topik yang bersifat kritis, dukungan, maupun desakan reformasi, penelitian ini diharapkan dapat memberikan kontribusi berupa *baseline* kuantitatif bagi DJBC dalam merumuskan strategi komunikasi publik yang lebih responsif dan transparan di era kepemimpinan yang baru.

## METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan seperti ditunjukkan pada Gambar 1.



Gambar 1 Alur Tahapan Penelitian

### Pengumpulan Data

Data dikumpulkan menggunakan teknik web scraping pada platform YouTube menggunakan library *youtube-comment-downloader* dalam bahasa Python. Web scraping digunakan dalam penelitian ini karena kemudahan fitur untuk mengumpulkan data relevan yang jumlahnya banyak dari berbagai sumber, sehingga sifat data aktual dan kredibilitasnya tidak diragukan (Alfarizi et al., 2026). Sebanyak 15 video pemberitaan relevan diseleksi berdasarkan kata kunci: "bea cukai viral 2025", "menkeu purbaya bea cukai", "djbc penindakan", "bea cukai kena tegur", dan "viral bea cukai terbaru 2026". Dari video-video tersebut, diperoleh 937 komentar publik pada periode 8 September 2025 hingga 18 April 2026.

### Preprocessing Data

Tahapan preprocessing dilakukan dalam urutan berikut (Karas et al., 2022) (Syahputra Novelan & Iqbal, 2025):

1. **Case Folding**  
Tahapan ini mengubah semua huruf menjadi huruf kecil agar data konsistensi dan tidak menambah variasi nilai akibat perbedaan huruf.
2. **Cleaning Data**  
Tahap ini dilakukan proses penghapusan URL, mention, hashtag, angka, dan karakter non-ASCII dan penghapusan multiple spasi.
3. **Normalization**  
Proses dilakukan perbaikan penulisan kata – kata yang memiliki bentuk tidak baku atau format penulisan yang unik/berbeda. Contoh: Tdk = tidak, menkeu =

- menteri keuangan, impor = import, bnar = benar.
4. Tokenization  
Komentar yang sudah dibersihkan tersebut kemudian dipecah menjadi kata per kata agar mempermudah algoritma melakukan analisis dan memodelkan topik (Mramra & Nataliani, 2025).
  5. Stopword Removal  
Tahap ini menggunakan library Sastrawi dan custom stopwords untuk menghapus kata-kata informal (yg, sih, aja, bgt, dll).
  6. Stemming  
Tahapan ini menggunakan StemmerFactory dari Sastrawi, gunanya untuk mengembalikan kata ke bentuk dasarnya agar mengurangi variasi kata supaya stabil untuk dimodelkan (Muhammad Ramli & Indah Ratnasari, 2025).
  7. Filter dokumen dengan minimal 3 kata setelah preprocessing. Dari 937 data awal, tersisa 832 dokumen yang siap untuk pemodelan.

### Pembentukan Dictionary dan Corpus

Seluruh dokumen hasil preprocessing dikonversi menjadi Bag-of-Words (BoW) menggunakan library Gensim (Kornelius & Surya, 2026). Dictionary difilter dengan parameter  $no\_below = 5$  (kata harus muncul di minimal 5 dokumen) dan  $no\_above = 0.5$  (kata tidak boleh muncul di lebih dari 50% dokumen) untuk menghilangkan kata yang terlalu jarang atau terlalu umum. Hasil akhir: 404 kata unik dalam dictionary. Sebelum pembentukan dictionary, dilakukan deteksi bigram menggunakan Phrases (Gensim) untuk menggabungkan frasa seperti "bea\_cukai" dan "pak\_purbaya" sebagai satu token.

### Pemodelan Latent Dirichlet Allocation (LDA)

Latent dirichlet allocation (LDA) merupakan algoritma unsupervised learning yang bertujuan untuk mengelompokkan kata menjadi topik-topik utama yang tersembunyi di dalam

sekumpulan dokumen teks. Pemodelan dilakukan menggunakan gensim.models.LdaModel dengan parameter:  $passes = 10$  untuk tuning dan  $passes = 20$  untuk model final,  $iterations = 100-200$ ,  $alpha = 'auto'$ ,  $eta = 'auto'$ , dan  $random\_state = 42$  untuk reproduktibilitas.

### Evaluasi dan Penentuan K Optimal

Evaluasi digunakan untuk menilai bagaimana kualitas topik yang dihasilkan algoritma LDA. Penelitian ini menggunakan *coherence score* (Skor Koherensi) untuk mengukur hubungan semantik antara kata-kata terbanyak dalam sebuah topik dan menilai seberapa masuk akal sebuah topik (Nursyahrina et al., 2024) (Asyifa & Novelan, n.d.). *Perplexity log* digunakan untuk menilai kinerja model jika dimasukkan data baru (diluar data latih) (Abdurrazzaq, 2023).

Untuk menilai Pencarian jumlah topik optimal dilakukan dengan melatih 14 model LDA pada rentang  $K = 2$  hingga  $K = 15$ . Setiap model dievaluasi dengan Coherence Score ( $C_v$ ) menggunakan CoherenceModel Gensim dan Log Perplexity. Nilai K dengan Coherence tertinggi dipilih sebagai model final.

## HASIL DAN PEMBAHASAN

### Statistik Deskriptif Data

Tabel 1 Statistik Deskriptif Dataset

Parameter	Nilai
Total komentar mentah	942
Setelah deduplikasi & filter	858
Setelah preprocessing NLP	832
Ukuran dictionary (kata unik)	404
Periode data	8 Sep 2025 – 18 Apr 2026
Sumber data utama	YouTube Comments

**Tuning Jumlah Topik Optimal**

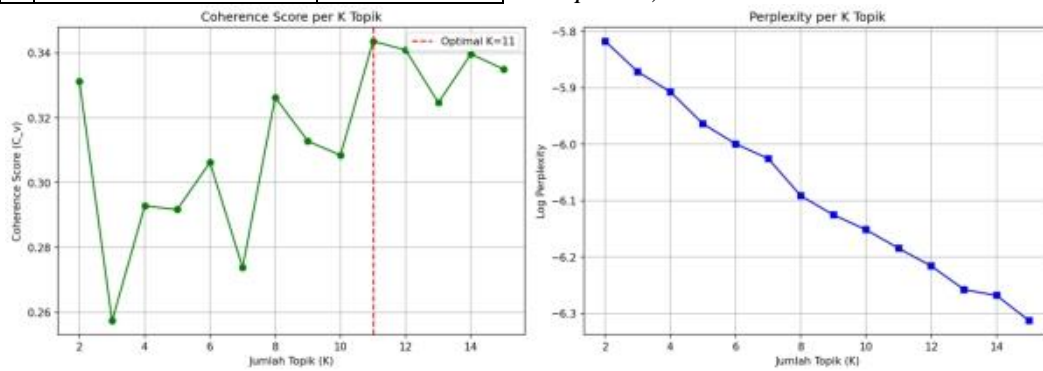
Pencarian nilai K optimal dilakukan dengan melatih model LDA pada K = 2 hingga K = 15 dan mengevaluasi Coherence Score dan Perplexity masing-masing. Hasil selengkapnya ditampilkan pada Tabel 2 dan Gambar 2.

**Tabel 2 Hasil Tuning Coherence Score dan Perplexity**

K	Coherence Score (C_v)	Log Perplexity
2	0,3312	-5,8174
3	0,2574	-5,8715

4	0,2928	-5,9072
5	0,2917	-5,9639
6	0,3063	-5,9994
7	0,2738	-6,0257
8	0,3263	-6,0917
9	0,3128	-6,1254
10	0,3085	-6,1519
11	0,3436 *	-6,1849
12	0,3409	-6,2163
13	0,3247	-6,2579
14	0,3397	-6,2684
15	0,3350	-6,3133

Keterangan: \* = Coherence tertinggi (K optimal)



**Gambar 2 Grafik Coherence Score dan Perplexity per K Topik**

Berdasarkan Tabel 2 dan Gambar 2, dapat diamati bahwa Coherence Score mencapai nilai tertinggi pada K = 11 dengan skor 0,3436. Meskipun Perplexity terus menurun seiring bertambahnya K (yang mengindikasikan model semakin fit) (Sofa et al., 2025), penelitian Chang et al. [3] menegaskan bahwa Coherence lebih relevan untuk interpretabilitas

manusia. Oleh karena itu, K = 11 dipilih sebagai jumlah topik optimal.

**Interpretasi 11 Topik**

Model final LDA dengan K = 11 menghasilkan topik-topik dengan keyword dominan seperti ditampilkan pada Tabel 3. Label topik diberikan secara manual berdasarkan analisis semantik atas 15 kata teratas per topik.

**Tabel 3 Hasil 11 Topik LDA dengan Label Interpretasi**

No	Label Topik	Top Keywords	Jml Doc
1	Tuntutan Hukuman Berat untuk Oknum	pak, bea_cukai, biar, masuk, hukum_mati, indonesia, barang_ilegal, miskin, dpr, menteri	80
2	Dukungan Publik kepada Menkeu Purbaya	purbaya, baik, bea_cukai, tri, baru, pak, takut, jangnan, pemerintah, dukung	97
3	Desakan Pembersihan Pegawai Bea Cukai	mau, sdh, orang, kalau, kerja, sama, jabat, dulu, pecat, benah, tangkap	77
4	Persepsi Negatif: "Sarang Maling"	bea_cukai, maling, semua, jadi, jabat, termul, ganti, sgs, parah, sarang, jahat	99
5	Ketimpangan Pegawai	banyak, tdk, kerja, negara, apa, rakyat, jabat,	89

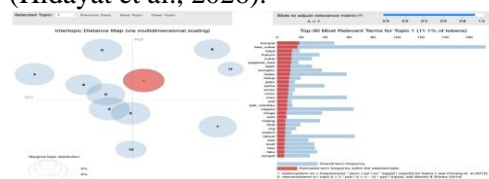
	vs Rakyat	korupsi, indonesia, rumah, cari	
6	Isu Korupsi Pegawai Bea Cukai	korupsi, bea_cukai, kaya, hukum, cukai, pegawai_bea, koruptor, hidup, jalan, aman	66
7	Barang Ilegal & Penyelundupan	barang, pak_pur, ganti, tugas, terus, uang, ilegal, tahu, jgn, biasa, usaha, lama	54
8	Cukai Rokok & "Tikus" di Instansi	pak, rokok, cuma, bea_cukai, semua, beku, awas, tikus, begini, utk, selesai, atur	44
9	Doa & Harapan Publik kepada Menkeu	bea_cukai, pak_purbaya, moga, selalu, allah, oknum, lindung, barang, sehat, negara, kpk	72
10	Apresiasi Ketegasan Menkeu	pak_purbaya, bea_cukai, baru, kerja, pecat, smua, berani, kpk, tegas, kasih, langsung	80
11	Politik: Prabowo-Purbaya vs Koruptor	rakyat, koruptor, pak, lama, negara, tahun, semua, hukum, kena, laku, manusia, prabowo	74

Berdasarkan Tabel 3, topik-topik yang dihasilkan dapat dikelompokkan ke dalam tiga kategori besar. Pertama, topik bernada kritis (Topik 1, 4, 5, 6, 7, 8) yang mencakup tuntutan hukuman berat, persepsi negatif terhadap institusi, isu korupsi, dan penyelundupan barang ilegal. Kelompok ini mendominasi dengan total 432 dokumen atau 52,3% dari seluruh korpus.

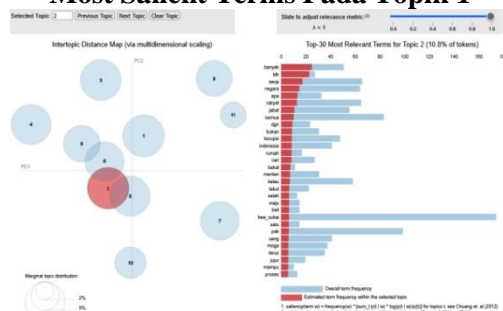
Kedua, topik bersifat dukungan (Topik 2, 9, 10) yang merepresentasikan apresiasi publik terhadap Menteri Keuangan Purbaya, doa untuk keberhasilan, dan pujian atas ketegasan kebijakan. Kelompok ini mencakup 249 dokumen atau 29,9%. Ketiga, topik desakan reformasi (Topik 3, 11) yang fokus pada tuntutan pembersihan pegawai dan reformasi kelembagaan. Kelompok ini sebanyak 151 dokumen atau 18,1%.

**Analisis Intertopic Distance Map**

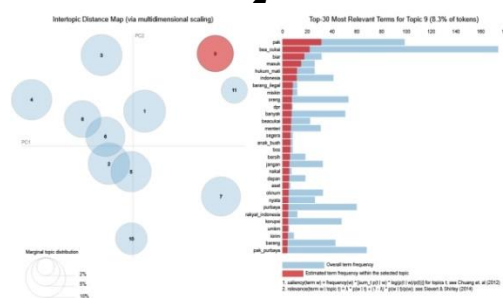
Visualisasi pyLDAvis (Gambar 3) menampilkan peta jarak antar-topik berdasarkan teknik multidimensional scaling. Ukuran lingkaran merepresentasikan prevalensi topik dalam korpus, sementara jarak antar-lingkaran menunjukkan kemiripan semantik (Hidayat et al., 2026).



**Gambar 3 Visualisasi pyLDAvis Intertopic Distance Map dan Top-30 Most Salient Terms Pada Topik 1**



**Gambar 4 Visualisasi pyLDAvis Topik 2**



**Gambar 5 Visualisasi pyLDAvis Topik 3**

Dari Gambar tersebut, terlihat bahwa Topik 1, 2, 5, 6, dan 8 berada dalam cluster yang berdekatan, mengindikasikan kesamaan semantik dalam isu kritik dan korupsi. Topik 3, 9, dan 11 terpisah cukup jauh dari cluster lain, menunjukkan bahwa topik-topik tersebut unik dan spesifik. Topik 4 merupakan topik terbesar dengan 99

dokumen, konsisten dengan ukuran lingkaran terbesar pada visualisasi.

Kata dengan saliency tertinggi adalah "bea\_cukai" (frekuensi >170), diikuti "pak" (>100) dan "pak\_purbaya" (>75). Tingginya frekuensi istilah-istilah ini memvalidasi bahwa korpus yang dikumpulkan telah relevan dan fokus pada topik penelitian.

### Implikasi Hasil

Dominasi topik kritis (52,3%) menunjukkan bahwa publik Indonesia pada periode penelitian memiliki tingkat kepercayaan yang rendah terhadap integritas aparatur Bea Cukai (Siahaan, 2025). Namun, kehadiran topik dukungan (29,9%) mengindikasikan adanya momentum positif seiring dengan kepemimpinan baru Menkeu Purbaya. Topik desakan reformasi (18,1%) menjadi sinyal kuat bahwa masyarakat mengharapkan tindakan konkret, bukan sekadar pernyataan retorik.

Temuan ini memiliki implikasi praktis bagi DJBC: (1) perlunya strategi komunikasi publik yang lebih responsif terhadap isu-isu spesifik yang teridentifikasi; (2) pentingnya transparansi dalam proses penindakan oknum untuk membangun kembali kepercayaan publik; (3) perlunya pemanfaatan data media sosial secara sistematis sebagai bagian dari evaluasi kebijakan.

### SIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma Latent Dirichlet Allocation (LDA) untuk mengidentifikasi topik dominan pemberitaan Bea Cukai di media sosial pasca pergantian Menteri Keuangan pada September 2025. Dari 832 dokumen hasil preprocessing, model LDA optimal diperoleh pada  $K = 11$  dengan Coherence Score tertinggi sebesar 0,3436. Sebelas topik yang dihasilkan mencerminkan spektrum opini publik yang mencakup kritik terhadap integritas institusi

(52,3%), dukungan terhadap kepemimpinan Menkeu Purbaya (29,9%), dan desakan reformasi kelembagaan (18,1%). Kontribusi penelitian ini meliputi: (1) penerapan metodologi LDA pada domain spesifik pemberitaan Bea Cukai yang belum banyak diteliti sebelumnya; (2) penyediaan baseline pemahaman kuantitatif atas persepsi publik terhadap DJBC di era baru Kemenkeu; (3) demonstrasi praktis penggunaan Coherence Score  $C_v$  sebagai metrik evaluasi utama dibandingkan Perplexity.

Saran untuk penelitian selanjutnya: (1) perluasan sumber data ke platform lain seperti Twitter/X dan TikTok untuk memperoleh representasi yang lebih komprehensif; (2) integrasi analisis sentimen dengan LDA (seperti ASUM atau JST model) untuk memperoleh dimensi tambahan; (3) perbandingan LDA dengan model topic modeling modern berbasis transformer seperti BERTopic atau Top2Vec; (4) analisis temporal untuk melihat perubahan topik sepanjang masa jabatan Menkeu Purbaya.

### DAFTAR PUSTAKA

- Abdurrazzaq, M. A. (2023). Analisis Ulasan Aplikasi MyPertamina Menggunakan Topic Modeling dengan Latent Dirichlet Allocation. *KALBISCIENTIA Jurnal Sains Dan Teknologi*, 10(1), 1–6. <https://doi.org/10.53008/kalbiscientia.v10i1.694>
- Alfarizi, N., Lydia, P., Novelan, M. S., Putra, A., & Sinurat, S. (2026). *Comparative Machine Learning Analysis for Sentiment Classification of Sumatra Disaster 2025*. 3(1), 68–76.
- Asyifa, N., & Novelan, M. S. (n.d.). *Performance Analysis of the K-Means Algorithm in Classifying Organic and Inorganic Waste Types*. 6681–6690.
- Farhan, A., Ridat, I., & Isnain, A. R. (2025). *Pemodelan topik dengan lda*

- untuk memahami persepsi publik terhadap apple vision pro melalui analisis berita twitter. *10(4)*, 3233–3246.
- Farta Wijaya, R., Sitorus, Z., & Diva, K. (2025). Krisna Diva Analisis Sentimen Publik Berbasis KNN terhadap Kinerja Purbaya dan Sri Mulyani. *Jurnal Pustaka AI*, *5(3)*, 528–537. <https://doi.org/10.55382/jurnalpustakaaai.v5i3.1461>
- Hidayat, F. M., Sanjaya, H., Purnomo, D., Wiranto, H., Yayasan, U., Imam, P., & Majalengka, B. (2026). *PEMODELAN TOPIK BERITA NASIONAL INDONESIA MENGGUNAKAN*. *12(1)*, 33–39.
- Karas, B., Qu, S., Xu, Y., & Zhu, Q. (2022). Experiments with LDA and Top2Vec for embedded topic discovery on social media data—A case study of cystic fibrosis. *Frontiers in Artificial Intelligence*, *5*. <https://doi.org/10.3389/frai.2022.948313>
- Kornelius, Y., & Surya, I. G. (2026). Analisis Sentimen Gemini AI Menggunakan Multinomial Naïve Bayes dengan TF-IDF dan BoW. *4*, 253–264.
- Mramra, W. Y. Y., & Nataliani, Y. (2025). Penentuan bidang unggulan akademik universitas melalui metode topik Linear Discriminant Analysis (LDA). *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, *4(1)*, 93–105. <https://doi.org/10.24246/itexplore.v4i1.2025.pp93-105>
- Muhammad Ramli, R., & Indah Ratnasari, C. (2025). Pemodelan Topik pada Komunitas Ekspresi Emosi Negatif di Media Sosial X Menggunakan LDA. *Media Online*, *6(1)*, 224–234. <https://doi.org/10.47065/bulletinsr.v6i1.877>
- Nursyahrina, Defit, S., & Sovia, R. (2024). Metode BERTopic dan LDA untuk Analisis Tren Penelitian Bidang Ilmu Komputer. *Jurnal KomtekInfo*, *11(4)*, 332–341. <https://doi.org/10.35134/komtekinfo.v11i4.580>
- 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. *Journal of Information Engineering and Educational Technology*, *8(1)*, 14–17. <https://doi.org/10.26740/jieet.v8n1.p14-17>
- Salwa, G. (2025). EduTIK: Jurnal Pendidikan Teknologi Informasi dan Komunikasi Volume 1 Nomor 5, Oktober 2021. *EduTIK: Jurnal Pendidikan Teknologi Informasi Dan Komunikasi*, *2(5)*, 773.
- Sepriadi, H., Sekarsari, C. R., Iriany, A., Achmad, A., Fernandes, R., Brawijaya, U., Korespondensi, P., Society, C., Allocation, L. D., & Modeling, T. (2025). Kombinasi Analisis Bibliometrik Dengan Latent Dirichlet Combination of Bibliometric Analysis With Latent Dirichlet Allocation As Topic Modeling of Cashless Society Behavior. *12(2)*, 331–338. <https://doi.org/10.25126/jtiik.2025129244>
- Septiara, D. I., Arifianto, D., & Suharso, W. (2026). Analisis Topik Pada Transkrip Video Pidato Politik Menggunakan Metode Latent Dirichlet Allocation ( LDA ) Topic Analysis in Political Speech Video Transcripts Using the Latent Dirichlet Allocation ( LDA ) Method. *11(1)*, 1–10.
- Siahaan, A. L. (2025). Pengaruh Integritas dan Kapabilitas Kerja terhadap Kinerja Pegawai melalui Profesionalisme. *Jurnal Syntax Admiration*, *6(1)*, 431–450. <https://doi.org/10.46799/jsa.v6i1.2049>
- Sofa, N., Utomo, F. S., & Saputro, R. E. (2025). Eksplorasi Model Hybrid Transformer-Latent Semantic Analysis ( LSA ) Untuk Pemahaman

- 
- Konteks Teks Berita Berbahasa Indonesia Fakultas Ilmu Komputer , Universitas Amikom Purwokerto , Indonesia Exploration of Hybrid Transformer Model-Latent Semantic Analsi. 5(5), 1239–1252.*
- Syahputra Novelan, M., & Iqbal, M. (2025). Comparative Analysis of the Performance of K-Nearest Neighbor (K-NN) and Naive Bayes Algorithms on User Satisfaction Levels of the Tokopedia Application. *Proceedings of The International Conference on Computer Science, Engineering, Social Science, and Multi-Disciplinary Studies, 1*, 218–224.
- Wakhid, N. K., Rosyid, H. Al, Ketintang, J., Gedung, W., & Timur, J. (2026). *ANALISIS PERSEPSI PUBLIK TERHADAP RESHUFFLE MENTERI KEUANGAN 2025 BERDASARKAN TOPIC MODELING LDA PADA KOMENTAR YOUTUBE. 10(1), 646–654.*
- Yahya, S., Sitorus, Z., Iqbal, M., Nasution, D., & ... (2025). Analisis Sentimen Penerapan Deep Learning dan Analisis Sentimen terhadap Gap Kompetensi Lulusan Lembaga Pendidikan dan Pelatihan Vokasi terhadap Dunia .... *Bulletin of Information ...*, 6(2), 0–11. <https://doi.org/10.47065/bit.v5i2.202>