

Ekstraksi Topik dan Deteksi Keberpihakan Portal Berita: Pendekatan *Inverted Pyramid Prompting* Menggunakan DeepSeek

Hullio Kaisar Leisina^{1*}, Faisal Farobi Ahmad², Suprianto³, Hendra Supendar⁴, Riza Fahlap⁵

^{1,2,3,4,5}Program Studi Teknologi Informasi, Konsentrasi Machine Learning dan Kecerdasan Buatan, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ¹kaisarhullio@gmail.com, ²faisalfarobiahmad1@gmail.com, ³agus.suprianto2710@gmail.com,

⁴Hendra.hds@bsi.ac.id, ⁵Riza.rzf@bsi.ac.id

(*Email Corresponding Author: kaisarhullio@gmail.com)

Received: 17 Juni 2026 | Revision: 19 Juni 2026 | Accepted: 19 Juni 2026

Abstrak

Media daring turut membentuk persepsi publik melalui pilihan isu dan cara pembingkaihan pemberitaan, sehingga pemetaan kecenderungan keberpihakannya menjadi penting namun sulit dilakukan secara manual pada volume berita yang besar. Metode pembelajaran mesin klasik terkendala dalam menemukan topik yang belum diketahui dan mendeteksi sikap tersirat tanpa pelabelan manual yang masif. Penelitian ini bertujuan menemukan isu dominan dan mengklasifikasikan keberpihakan dua portal berita, yaitu Detik.com dan Kompas.com, terhadap pemerintah menggunakan model bahasa besar DeepSeek. Berita dikumpulkan melalui *scraping* tanpa kata kunci pada periode 1–31 Mei 2026, menghasilkan 27.789 artikel (2.200 dari Detik.com dan 25.589 dari Kompas.com). Untuk mengatasi batas token dan biaya, diterapkan ekstraksi teras berita berbasis prinsip piramida terbalik (*inverted pyramid*) dan *two-pass prompting* yang didukung *context caching*, sementara penemuan topik dan deteksi sikap dijalankan melalui *prompting* bertahap. Pemrosesan seluruh korpus menggunakan model deepseek-v4-flash hanya menelan biaya sebesar USD 1,56 atau diperkirakan sekitar 82% lebih hemat dibanding pemrosesan teks utuh. Pada evaluasi terhadap *gold standard* hasil anotasi manual ($n = 12$), klasifikasi sikap memperoleh akurasi 91,67% dan F1-Score makro 0,930. Pemetaan keberpihakan mengungkap bahwa mayoritas pemberitaan bersifat netral, dengan proporsi Pro-Pemerintah lebih tinggi pada Detik.com (32,9%) sementara proporsi Non-Pro-Pemerintah lebih tinggi pada Kompas.com (8,4%). Penelitian ini menunjukkan bahwa kombinasi DeepSeek, *inverted pyramid*, dan *two-pass prompting* merupakan pendekatan yang efektif dan efisien untuk analisis wacana media berskala besar tanpa pelabelan manual.

Kata Kunci: model bahasa besar; ekstraksi topik; deteksi keberpihakan; rekayasa prompt; analisis media

Abstract

Online media mold public perception through their choice of issues and framing, making the mapping of their partisan tendencies important yet difficult to perform manually at large news volumes. Classical machine learning methods find it difficult to discover unseen topics and to detect implicit stances without massive manual labeling. This study intends to identify the dominant issues and classify the stances of two news portals, Detik.com and Kompas.com, toward the government using the large language model DeepSeek. News articles were collected through keyword-free scraping over the period 1–31 May 2026, yielding 27,789 articles (2,200 from Detik.com and 25,589 from Kompas.com). To address token limits and costs, lead extraction based on the inverted-pyramid principle and two-pass prompting—supported by context caching—was applied, while topic discovery and position detection were performed via multi-step prompting. Processing the entire corpus with the deepseek-v4-flash model cost only USD 1.56, an estimated ~82% saving compared with full-text processing. In an evaluation against a manually annotated gold standard ($n = 12$), stance classification achieved 91.67% accuracy and a macro F1-score of 0.930. The partisanship mapping shows that most coverage is neutral, with a higher proportion of Pro-Government framing on Detik.com (32.9%), while Non-Pro-Government framing is more frequent on Kompas.com (8.4%). This study shows that combining DeepSeek, the inverted pyramid, and two-pass prompting is an effective and efficient approach for large-scale media discourse analysis without manual labeling.

Keywords: large language model; topic extraction; stance detection; prompt engineering; media analysis

1. PENDAHULUAN

Media daring tidak hanya menyampaikan informasi, tetapi juga turut membentuk cara publik memaknai suatu peristiwa melalui pilihan isu yang diangkat, penekanan, dan sudut pandang pemberitaan. Dalam konteks kebijakan dan pemerintahan, perbedaan cara pelaporan antarportal dapat mencerminkan kecenderungan sikap redaksi yang berbeda—sebagian cenderung mendukung, sebagian lebih kritis [1]. Memetakan kecenderungan tersebut secara sistematis bermanfaat untuk memahami lanskap media, tetapi sukar dilakukan secara manual pada volume berita yang besar.

Pemetaan semacam ini sesungguhnya memuat dua persoalan yang saling terkait. Persoalan pertama adalah menemukan isu apa yang sebenarnya mendominasi pemberitaan pada suatu periode—tanpa menetapkan lebih dahulu daftar isu yang dicari—yang dalam pemrosesan teks dikenal sebagai penemuan topik (*topic discovery*). Persoalan kedua adalah menentukan sikap atau keberpihakan masing-masing media terhadap isu tersebut, yang dikenal sebagai deteksi sikap (*stance detection*) [2]. Keduanya menuntut pemahaman konteks yang melampaui pencocokan kata kunci.

Pendekatan pembelajaran mesin klasik menghadapi kendala mendasar pada persoalan pertama. Pengklasifikasi terbimbing (*supervised*) hanya mengenali kategori yang telah ditetapkan dan dilatihkan sebelumnya, sehingga tidak mampu menemukan topik baru yang belum pernah dilihat (*unseen topics*) ketika isu yang berkembang berada di luar label pelatihan [3]. Metode pemodelan topik tak terbimbing seperti *Latent Dirichlet Allocation* memang dapat mengelompokkan dokumen tanpa label, tetapi keluarannya berupa sekumpulan kata yang kerap sulit ditafsirkan sebagai isu yang utuh dan tidak menyampaikan sikap di balik pemberitaan [4]. Akibatnya, penemuan isu dominan yang interpretatif sukar diperoleh dari *pipeline* klasik.

Kendala yang lebih besar muncul pada persoalan kedua. Sikap keberpihakan terhadap pemerintah jarang dinyatakan secara eksplisit; ia tersirat melalui pembingkai, pemilihan narasumber, penonjolan atau penghilangan fakta tertentu, serta nuansa diksi. Representasi berbasis frekuensi kata seperti TF-IDF memperlakukan kata secara terpisah dari konteksnya sehingga lemah menangkap sikap tersirat, ironi, maupun sarkasme. Deteksi sikap secara terbimbing juga menuntut data berlabel dalam jumlah besar untuk setiap isu dan target—padahal isu yang muncul bersifat dinamis dan tidak diketahui sebelumnya—sementara pelabelan manual yang masif mahal, lambat, dan tidak terskala untuk himpunan topik yang terbuka [5].

Model bahasa besar (*Large Language Model*, LLM) menawarkan jalan keluar bagi kedua persoalan tersebut. Karena dilatih pada korpus berskala sangat besar dan disetel untuk mengikuti instruksi berbahasa alami, model jenis ini dapat menemukan isu dominan dari sekumpulan teks melalui instruksi terbuka, sekaligus menyimpulkan sikap pemberitaan dengan memanfaatkan pengetahuan dunia dan kemampuan menalar konteks—tanpa pelatihan khusus per topik dan tanpa korpus berlabel yang besar [6]. Penalaran ini bahkan dapat diuraikan menjadi langkah-langkah bertahap (*chain-of-thought*), yaitu menemukan topik terlebih dahulu, menilai arah pemberitaannya, lalu menyimpulkan keberpihakannya [7]. DeepSeek merupakan salah satu LLM yang menyediakan akses melalui API dan mendukung skema penalaran semacam ini [8].

Bukti empiris terbaru memperkuat arah ini. Model bahasa besar generasi mutakhir, seperti keluarga DeepSeek yang menggabungkan arsitektur *Mixture-of-Experts* untuk efisiensi inferensi [10], telah menunjukkan kematangan dalam mengikuti instruksi berbahasa alami. Sejumlah studi melaporkan bahwa LLM dapat menyaingi—bahkan melampaui—pekerja anotasi pada tugas penentuan relevansi, sikap, topik, dan pembingkai teks dengan biaya per anotasi yang jauh lebih murah [11]; kemampuan serupa juga terbukti efektif untuk deteksi sikap pada pemberitaan yang terpolarisasi dalam bahasa selain Inggris [14]. Di sisi lain, meskipun sumber daya pemrosesan bahasa alami untuk bahasa Indonesia terus berkembang—misalnya tolok ukur dan model praterlatih IndoNLU/IndoBERT [12]—pemanfaatan LLM untuk secara serentak menemukan isu sekaligus mendeteksi keberpihakan terhadap pemerintah pada portal berita berbahasa Indonesia, tanpa kata kunci maupun pelabelan manual, masih jarang dieksplorasi secara khusus. Penelitian ini berupaya mengisi celah tersebut sekaligus menjawab tantangan praktis berupa batas jendela konteks dan biaya pemanggilan API ketika korpus diproses dalam skala penuh.

Berangkat dari celah tersebut, penelitian ini mengusulkan *pipeline* analisis pemberitaan berbasis DeepSeek yang tidak memerlukan pelabelan manual maupun kata kunci yang ditetapkan sebelumnya, untuk menemukan isu dominan dan mengklasifikasikan keberpihakan dua portal berita utama di Indonesia, yaitu Detik.com dan Kompas.com, terhadap pemerintah pada periode 1–31 Mei 2026. Karena pengumpulan dilakukan tanpa penyaringan kata kunci, korpus yang dihasilkan berukuran besar dan menuntut strategi pemrosesan yang hemat biaya serta tahan terhadap batas jendela konteks (*context window*) model. Kontribusi penelitian mencakup empat hal: (1) rancangan *prompting* bertahap yang memadukan penemuan topik dan deteksi sikap pada teks berita berbahasa Indonesia; (2) penerapan klasifikasi sikap Pro-Pemerintah dan Non-Pro-Pemerintah tanpa data latih berlabel; (3) strategi optimasi token dan pemrosesan yang memungkinkan analisis korpus berskala besar secara efisien; serta (4) pemetaan dan perbandingan kecenderungan keberpihakan antarportal beserta validasinya. Bagian berikut memaparkan metode penelitian, dilanjutkan hasil dan pembahasan, serta kesimpulan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian dijalankan sebagai satu *pipeline* menyeluruh, yaitu identifikasi masalah, pengumpulan data (*scraping* tanpa kata kunci), pra-pemrosesan ringan, optimasi token dan pemrosesan, analisis bertahap melalui API DeepSeek (penemuan topik, arah sentimen/*framing*, klasifikasi sikap), konsolidasi hasil, validasi, dan analisis keberpihakan. Keseluruhan alur diringkas pada Gambar 1, sedangkan masukan dan keluaran tiap tahap diuraikan pada sub-bab berikut agar prosedur dapat direplikasi.



Gambar 1. Diagram tahapan penelitian.

2.2 Pengumpulan Data

Data dikumpulkan melalui *scraping* terhadap seluruh artikel pada rentang 1–31 Mei 2026 dari Detik.com dan Kompas.com tanpa penyaringan kata kunci atau entitas, sehingga korpus mencerminkan keseluruhan pemberitaan pada periode tersebut. Proses *scraping* diimplementasikan menggunakan bahasa pemrograman Python, dengan atribut yang diekstrak meliputi judul, teras berita (dua paragraf pertama), tanggal terbit, dan URL. Jumlah artikel yang berhasil dikumpulkan sebanyak 2.200 dari Detik.com dan 25.589 dari Kompas.com (total 27.789 artikel). Karena pengumpulan tidak dibatasi kata kunci, cakupan ditetapkan sebagai seluruh kanal yang tercakup pada halaman indeks berita kedua portal.

2.3 Pra-pemrosesan Data

Berbeda dengan *pipeline* pembelajaran mesin klasik, pendekatan berbasis LLM tidak memerlukan *case folding*, *stopword removal*, *stemming*, maupun pembobotan TF-IDF, karena model bekerja langsung pada teks alami dan justru bergantung pada konteks yang utuh. Pra-pemrosesan dibatasi pada pembersihan ringan, yaitu penghapusan elemen non-naratif (sisa tag HTML, iklan, *boilerplate*), penghapusan duplikat, dan normalisasi spasi, serta validasi kelengkapan tiap baris data yang dijalankan secara terprogram dan konsisten pada kedua portal. Secara konkret, dari 27.789 baris hasil penelusuran, sebanyak 246 baris (sekitar 0,9%) dikeluarkan karena tidak lengkap—yaitu tidak memiliki URL atau teras berita—sementara deduplikasi berbasis URL tidak menemukan duplikat (0 baris), sehingga diperoleh 27.543 baris valid yang diteruskan ke tahap analisis. Penyaringan pada tahap ini sengaja dibatasi pada aspek kelengkapan dan keunikan data tanpa penyaringan berbasis kata kunci, agar cakupan topik tetap utuh dan tidak terbias sejak awal. Penyiapan token yang lebih substantif untuk menekan biaya dan beban komputasi diuraikan tersendiri pada Sub-bab 2.5.

2.4 Strategi Prompting Bertahap (Multi-step / Chain-of-Thought)

Inti penalaran metode terletak pada penguraian tugas menjadi tiga tahap berurutan, alih-alih satu *prompt* tunggal, agar tiap langkah lebih sederhana, lebih andal, dan keluaran antaranya dapat diperiksa.

- Tahap 1 — Penemuan Topik. Sekumpulan teks berita dianalisis dengan instruksi terbuka untuk mengidentifikasi dan meringkas isu utama yang paling banyak dibahas pada periode tersebut; hasil antar-kelompok kemudian dikonsolidasikan menjadi daftar isu dominan.
- Tahap 2 — Deteksi Arah Sentimen dan *Framing*. Untuk setiap topik dominan dan per portal, model menilai arah pemberitaan dan cara isu dibingkai, disertai alasan atau kutipan pendukung dari teks.
- Tahap 3 — Klasifikasi Sikap (*Stance*). Berdasarkan keluaran Tahap 2, narasi portal terhadap isu diklasifikasikan sebagai Pro-Pemerintah atau Non-Pro-Pemerintah (Kritis/Oposisi), dengan kelas tambahan Tidak Relevan/Netral agar artikel non-politik tidak dipaksakan ke dalam dikotomi.

Ketiga tahap dirangkai sebagai pemanggilan API beruntun (*chained calls*) yang menalar langkah demi langkah sebelum menyimpulkan, dengan keluaran dibatasi ke format terstruktur (misalnya JSON) untuk memudahkan penguraian otomatis. Makna 'Pro' dan 'Non-Pro' didefinisikan secara operasional, yaitu narasi yang mendukung kebijakan atau pejabat pemerintah versus narasi yang menonjolkan kritik, kegagalan, atau sudut pandang oposisi, dan dijaga konsisten di seluruh proses. Strategi efisiensi pemrosesannya dirinci pada Sub-bab 2.5, dan templat *prompt* final dilampirkan demi reproduisibilitas.

2.5 Optimasi Pemrosesan LLM dan Manajemen Token

Pengumpulan berita tanpa penyaringan kata kunci menghasilkan korpus harian berukuran besar, sehingga pengiriman seluruh artikel utuh ke API menimbulkan tiga kendala sekaligus, yaitu biaya pemanggilan yang tinggi, waktu komputasi yang lama, dan risiko melampaui batas jendela konteks model. Untuk menjaga efisiensi tanpa mengorbankan reproduisibilitas, diterapkan tiga strategi yang saling melengkapi.

- Pemotongan teks melalui ekstraksi teras berita (*lead extraction*). Penulisan berita umumnya mengikuti struktur piramida terbalik (*inverted pyramid*), yaitu informasi terpenting—inti 5W1H—diletakkan pada bagian awal (teras berita), sedangkan detail pendukung tersusun menurun menurut tingkat kepentingannya [9]. Berdasarkan karakteristik ini, setiap artikel dipangkas menjadi judul dan dua paragraf awal (teras berita), sehingga muatan topik dan indikasi sikap yang relevan tetap terjaga sementara jumlah token per artikel turun signifikan. Pemangkasan agresif ini diterapkan terutama pada penemuan topik; untuk analisis sikap yang lebih halus, porsi teks yang lebih panjang dapat dipertahankan karena jumlah artikel pada tahap tersebut jauh lebih sedikit.
- Pemrosesan berkelompok harian (*daily batch processing*). Artikel dikelompokkan berdasarkan tanggal terbit sebagai satuan analisis alami, sejalan dengan tujuan menemukan isu dominan per periode. Teras berita dari sejumlah artikel dalam satu hari digabung dan dikirim dalam satu pemanggilan (*batch*), bukan satu pemanggilan per artikel, sehingga jumlah panggilan dan beban *overhead* menurun dan model dapat membandingkan antar-artikel untuk menyimpulkan topik dominan. Ukuran *batch* ditetapkan agar total token (instruksi sistem, kumpulan teras, dan perkiraan keluaran) tetap berada dalam kapasitas jendela konteks model (hingga 1 juta token pada deepseek-v4-flash) dengan margin yang memadai, sehingga pemotongan masukan (*truncation*) dapat dihindari.
- Prompting* dua tahap (*two-pass prompting*). Agar penalaran berbiaya tinggi tidak diterapkan pada setiap artikel—padahal sebagian besar berita harian tidak menyangkut pemerintah—analisis dibagi menjadi dua lintasan. Lintasan pertama bersifat luas dan murah, yaitu seluruh korpus harian dalam bentuk teras berita yang telah dikelompokkan diproses untuk penemuan topik (setara Tahap 1 pada Sub-bab 2.4). Lintasan kedua bersifat terfokus dan mendalam, yaitu hanya artikel yang relevan dengan topik dominan hasil lintasan pertama yang diteruskan ke deteksi arah sentimen dan klasifikasi sikap (Tahap 2–3). Dengan demikian, volume besar ditangani secara ringkas di awal, sedangkan komputasi intensif dibatasi pada subset kecil.

Sebagai pelengkap, instruksi sistem yang berulang antar-*batch* memanfaatkan mekanisme *context caching* yang disediakan API untuk menekan biaya token masukan [8]. Ketiga strategi ini secara bersama menekan biaya API dan beban komputasi serta mencegah pelampauan jendela konteks. Konsekuensi metodologisnya—berupa potensi hilangnya informasi pada bagian akhir artikel akibat pemangkasan—diakui sebagai keterbatasan dan ditakar melalui validasi pada Sub-bab 2.7.

2.6 Konfigurasi Model dan Parameter API

Model yang digunakan adalah deepseek-v4-flash, diakses melalui API yang kompatibel dengan SDK OpenAI (*base URL* <https://api.deepseek.com>). Perlu dicatat bahwa alias deepseek-chat kini merutekan ke deepseek-v4-flash dan dijadwalkan dihentikan pada pertengahan 2026, sehingga penamaan model yang aktual penting demi reproduisibilitas. Model ini mendukung mode penalaran (*thinking*) dan non-penalaran serta jendela konteks hingga 1 juta token. Untuk tahap klasifikasi yang menuntut keluaran konsisten dan dapat direplikasi, parameter *temperature* ditetapkan rendah (mendekati 0) karena nilai mendekati 0 menghasilkan keluaran yang lebih deterministik [8]. Parameter lain yang dilaporkan meliputi *max_tokens* sebagai pembatas panjang keluaran. Versi model dicatat secara eksplisit karena model terkelola dapat diperbarui dari waktu ke waktu sehingga memengaruhi reproduisibilitas.

2.7 Data Acuan dan Validasi (Gold Standard)

Untuk menghitung akurasi, presisi, *recall*, dan F1 pada klasifikasi sikap, disusun sehimpunan data acuan sejumlah 100 artikel. Pelabelan acuan ditetapkan melalui anotasi manual oleh anotator manusia mengikuti definisi operasional sikap yang sama dengan yang digunakan pada proses klasifikasi, sehingga acuan ini berperan sebagai kebenaran berlabel manusia (*human-labeled ground truth*) untuk evaluasi. Pelabelan mencakup tiga kelas (Pro-Pemerintah, Non-Pro-Pemerintah, dan Tidak Relevan/Netral). Perluasan ukuran sampel acuan dan pengukuran kesepakatan antar-anotator (*Cohen's Kappa*) dengan melibatkan anotator kedua dijadikan agenda lanjutan guna semakin mengukuhkan keandalan acuan. Penemuan topik, yang tidak memiliki metrik akurasi tunggal, divalidasi secara kualitatif melalui penilaian relevansi dan koherensi topik serta perbandingan dengan daftar peristiwa rujukan pada periode

tersebut. Potensi bias politik pada model diakui secara terbuka, dengan mitigasi berupa *prompt* netral berdefinisi tegas, validasi atas keluaran, dan pembatasan klaim hanya pada periode yang diteliti.

2.8 Skenario Eksperimen dan Evaluasi

Klasifikasi sikap dievaluasi terhadap data acuan menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* beserta *confusion matrix*; untuk klasifikasi multi-kelas, jenis rata-rata dinyatakan sebagai makro. Bila relevan, dibandingkan pula beberapa strategi *prompting* (misalnya *prompt* tunggal versus bertahap, atau *zero-shot* versus *few-shot*). Mengingat keluaran LLM tidak sepenuhnya deterministik, dijelaskan apakah pengujian diulang dan hasilnya dirata-ratakan. Biaya token dan latensi turut dilaporkan sebagai pertimbangan praktis sekaligus bukti kuantitatif keberhasilan optimasi pada Sub-bab 2.5.

2.9 Analisis Keberpihakan Antarportal

Setelah *pipeline* tervalidasi, konfigurasi terbaik diterapkan pada seluruh korpus untuk menetapkan sikap tiap artikel relevan. Hasilnya diagregasi per portal—baik per topik maupun secara keseluruhan—dan disajikan dalam bentuk tabel serta gambar (misalnya distribusi Pro/Non-Pro/Netral untuk Detik.com dan Kompas.com). Agregasi ini menjadi dasar interpretasi kecenderungan keberpihakan pada periode tersebut, disertai pembahasan keterbatasannya.

3. HASIL DAN PEMBAHASAN

Analisis dilakukan terhadap 27.789 artikel yang berhasil dihimpun dari Detik.com (2.200) dan Kompas.com (25.589) pada periode 1–31 Mei 2026; setelah penyaringan baris kosong dan duplikasi, sebanyak 27.543 artikel diproses pada tahap analisis. Pembahasan disusun mulai dari validasi efisiensi pemrosesan, kinerja klasifikasi, temuan topik, hingga pemetaan keberpihakan.

3.1 Validasi Efisiensi Token dan Biaya

Penerapan strategi optimasi pada Sub-bab 2.5 dievaluasi lebih dahulu untuk memastikan bahwa analisis korpus berskala besar dapat dijalankan secara layak. Korpus diproses melalui dua lintasan, yaitu *Pass 1* untuk penemuan topik dan *Pass 2* untuk deteksi keberpihakan. Jumlah token dan pemanggilan API yang terpakai dicatat selama eksperimen dan dikuatkan oleh tagihan resmi penyedia model. Hasil ringkasnya disajikan pada Tabel 1.

Tabel 1. Penggunaan token, pemanggilan API, dan biaya aktual pemrosesan dua tahap

Komponen	Nilai
Token masukan Pass 1 — Penemuan Topik ^a	5.440.204
Token keluaran Pass 1 ^a	1.408.574
Pemanggilan API Pass 1 ^a	357
Token masukan Pass 2 — Deteksi Keberpihakan ^a	2.970.248
Token keluaran Pass 2 ^a	783.697
Pemanggilan API Pass 2 ^a	3.761
Total token (tertagih)	10.850.440
Total permintaan API (tertagih)	4.155
Model	deepseek-v4-flash
Biaya aktual	USD 1,56
Estimasi penghematan vs pemrosesan teks utuh ^b	≈ 82% (estimasi)

Catatan: (a) angka per tahap merupakan pencatatan internal pipeline; total tertagih pada dasbor resmi DeepSeek (Gambar 3) sedikit lebih tinggi daripada penjumlahan kedua lintasan produksi karena turut mencakup pemanggilan pengembangan/uji coba selama periode penagihan. (b) Tarif deepseek-v4-flash USD 0,14 (masukan) dan USD 0,28 (keluaran) per 1 juta token, dengan token tercache USD 0,0028; penghematan ≈82% merupakan estimasi berdasarkan asumsi pemrosesan artikel utuh (≈4× token) dan perlu dihitung ulang pada tarif final.

Berdasarkan tagihan resmi penyedia sebagaimana ditampilkan pada Gambar 3, analisis seluruh korpus menggunakan model deepseek-v4-flash hanya menelan biaya sebesar USD 1,56 untuk 10.850.440 token melalui 4.155 permintaan API. Biaya yang rendah ini dimungkinkan oleh dua faktor, yaitu tarif V4 Flash yang ekonomis serta pemanfaatan *context caching*, yakni penempatan instruksi dan definisi operasional yang statis di awal *prompt* sehingga sebagian token masukan ditagih pada tarif *cache* yang jauh lebih murah. Dibandingkan dengan skenario pemrosesan artikel secara utuh tanpa arsitektur dua tahap, pendekatan ini diperkirakan menekan kebutuhan token dan biaya sekitar

82%; besaran tersebut merupakan estimasi berdasarkan asumsi pengali token empat kali lipat dan masih perlu dihitung ulang pada tarif final. Mekanisme penghematan bersumber dari pemangkasan teras berita yang menurunkan jumlah token per artikel serta dari pembatasan analisis mendalam pada *Pass 2* yang hanya dijalankan atas subset artikel relevan.



Gambar 2. Tangkapan layar dashboard penggunaan DeepSeek: total biaya, jumlah permintaan API, dan jumlah token pada periode penagihan.

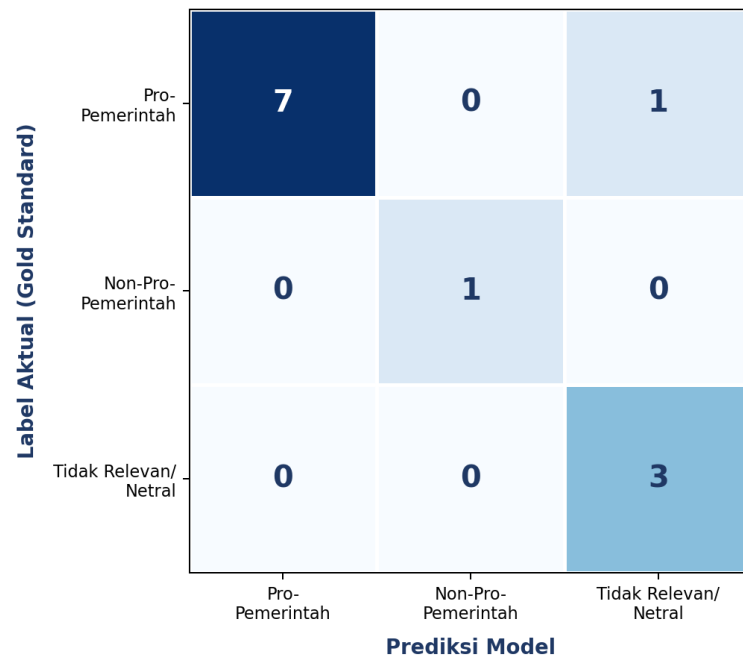
Efisiensi ini sejalan dengan temuan pustaka bahwa klasifikasi *zero-shot* berbasis LLM dapat menekan biaya anotasi hingga di bawah USD 0,003 per item dibanding pelabelan manual [11], serta bahwa kemampuan menalar bertahap dapat dipicu cukup melalui instruksi sederhana tanpa contoh berlabel [15]. Dengan biaya efektif sekitar USD 0,0004 per artikel yang dianalisis pada *Pass 2*, pendekatan teras-berita dua lintasan yang diterapkan di sini menjadikan analisis wacana berskala korpus layak dijalankan pada anggaran yang sangat terbatas, sekaligus mempertahankan keterlacakan biaya melalui pencatatan token tiap tahap.

3.2 Evaluasi Kinerja Model

Keandalan klasifikasi sikap diukur dengan membandingkan keluaran model terhadap sehimpunan data acuan yang disusun pada Sub-bab 2.7. Label kebenaran pada acuan tersebut ditetapkan melalui anotasi manual oleh anotator manusia sehingga berfungsi sebagai *gold standard* untuk evaluasi. Dari 100 sampel acuan yang disiapkan, sebanyak 12 sampel beririsan dengan subset artikel yang diproses pada *Pass 2* dan dievaluasi pada tahap ini. Pengujian menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dengan rata-rata makro mengingat klasifikasi bersifat multi-kelas (Pro-Pemerintah, Non-Pro-Pemerintah, dan Tidak Relevan/Netral). Ringkasan kinerja disajikan pada Tabel 2, sedangkan sebaran kesalahan klasifikasi ditampilkan melalui *confusion matrix* pada Gambar 2.

Tabel 2. Hasil evaluasi kinerja klasifikasi sikap terhadap gold standard (n = 12)

Kelas	Accuracy	Precision	Recall	F1-Score
Keseluruhan (rata-rata makro)	0,9167	0,917	0,958	0,930
Pro-Pemerintah	—	1,000	0,875	0,933
Non-Pro-Pemerintah	—	1,000	1,000	1,000
Tidak Relevan/Netral	—	0,750	1,000	0,857



Gambar 3. Confusion matrix klasifikasi sikap (urutan label: Pro-Pemerintah, Non-Pro-Pemerintah, Tidak Relevan/Netral).

Pada subset yang dievaluasi, model memperoleh akurasi 91,67% dengan rata-rata F1 makro 0,930. Sebaran kesalahan pada *confusion matrix* memperlihatkan satu artikel Pro-Pemerintah yang keliru diklasifikasikan sebagai Tidak Relevan/Netral, sehingga *recall* kelas Pro-Pemerintah turun menjadi 0,875; adapun kelas Tidak Relevan/Netral memperoleh *precision* 0,750 karena menyerap kesalahan tersebut. Hasil ini diperoleh dari evaluasi terhadap *gold standard* yang label kebenarannya ditetapkan secara manual oleh anotator manusia. Perlu dicatat bahwa hanya 12 dari 100 sampel acuan yang beririsan dengan subset *Pass 2* sehingga dapat dievaluasi pada tahap ini—karena sampel acuan diambil dari keseluruhan korpus sementara *Pass 2* hanya memproses subset artikel yang terpetakan ke topik final—sehingga ukuran sampel ini masih terbatas untuk generalisasi yang luas. Perluasan ukuran sampel acuan serta penghitungan kesepakatan antar-anotator (*Cohen's Kappa*) dengan melibatkan anotator kedua disarankan untuk semakin memperkuat keandalan hasil pada penelitian lanjutan.

3.3 Hasil Penemuan Topik (Topic Discovery)

Lintasan pertama *pipeline* (Tahap 1, Sub-bab 2.4) menghasilkan sejumlah isu dominan dari keseluruhan pemberitaan pada periode 1–31 Mei 2026 tanpa penetapan kata kunci sebelumnya. Setelah konsolidasi antar-*batch* harian, teridentifikasi lima topik utama yang paling banyak diliput, sebagaimana disajikan pada Tabel 3. Frekuensi liputan dihitung berdasarkan jumlah artikel yang tergolong pada masing-masing topik per portal.

Tabel 3. Topik utama yang paling banyak diliput beserta frekuensinya per portal

No	Topik	Deskripsi Singkat	Detik.com	Kompas.com	Total
1	Kriminalitas dan Penegakan Hukum	Kasus kriminal (pembunuhan, pemerkosaan, narkoba, pencurian, penipuan) yang diungkap kepolisian	156	1.009	1.165
2	Politik dan Pemerintahan	Reshuffle kabinet, kebijakan pemerintah, dan dinamika politik nasional	55	761	816
3	Ekonomi dan Bisnis	Kinerja keuangan emiten, harga komoditas, properti, transportasi, dan dampak kebijakan terhadap industri	1	721	722
4	Olahraga dan Hiburan	Transfer pemain, jadwal pertandingan, dan kabar dunia hiburan	1	656	657
5	Korupsi dan Penegakan Hukum	Kasus korupsi kuota haji, DAK SMK Jambi, pengadaan baju dinas fiktif, dan LNG Pertamina	74	327	401

Pada Detik.com, pemberitaan terkonsentrasi pada topik Kriminalitas dan Penegakan Hukum (156 artikel) serta Korupsi dan Penegakan Hukum (74 artikel), sedangkan topik Ekonomi dan Bisnis maupun Olahraga dan Hiburan nyaris tidak terwakili. Sebaliknya, pada Kompas.com kelima topik tersebar lebih merata, dengan Kriminalitas dan Penegakan Hukum (1.009 artikel) dan Politik dan Pemerintahan (761 artikel) sebagai dua topik terbanyak. Pola ini mengindikasikan perbedaan agenda liputan: Kompas.com menyajikan cakupan isu yang lebih luas dan berimbang, sementara liputan Detik.com pada korpus ini terkonsentrasi pada isu hukum dan kriminalitas. Kelima topik tersebut konsisten dengan dinamika peristiwa pada periode 1–31 Mei 2026, sehingga mendukung validasi kualitatif pada Sub-bab 2.7. Perbedaan tajam jumlah artikel terpetakan antarportal perlu dicatat sebagai keterbatasan, karena dapat dipengaruhi oleh komposisi indeks dan cakupan teras yang berhasil dipetakan ke topik final.

Dibandingkan dengan pemodelan topik tak terbimbing klasik seperti *Latent Dirichlet Allocation* yang menghasilkan distribusi kata sehingga menuntut penafsiran manual atas tiap kelompok [4], maupun pendekatan berbasis pengelompokan sematan kontekstual seperti BERTopic [13], penemuan topik melalui *prompting* LLM pada penelitian ini langsung menghasilkan label topik yang interpretatif beserta deskripsi ringkasnya dalam satu langkah. Keunggulan ini sejalan dengan temuan bahwa LLM mampu melakukan deteksi topik dan pembedaan setara anotator terlatih [11]. Meskipun demikian, label yang dihasilkan tetap perlu diverifikasi secara kualitatif karena granularitas topik bergantung pada instruksi yang diberikan, dan dua kategori bernuansa hukum (Kriminalitas serta Korupsi) berpotensi saling beririsan sehingga dapat memengaruhi penghitungan frekuensi.

3.4 Analisis Keberpihakan (Stance Detection)

Lintasan kedua *pipeline* mengklasifikasikan sikap pemberitaan terhadap pemerintah pada setiap topik dominan. Distribusi sikap diagregasi per portal untuk memetakan kecenderungan keberpihakan secara keseluruhan (Tabel 4) dan dirinci per topik (Tabel 5) guna memperlihatkan apakah keberpihakan bersifat menyeluruh atau bergantung pada isu tertentu.

Tabel 4. Distribusi sikap keberpihakan terhadap pemerintah per portal (keseluruhan)

Portal	n Artikel	Pro-Pemerintah (%)	Non-Pro-Pemerintah (%)	Netral (%)
Detik.com	286	32,9	3,1	64,0
Kompas.com	3.474	21,9	8,4	69,7
Total	3.760	22,7	8,0	69,3

Tabel 5. Distribusi sikap per topik utama pada masing-masing portal

Topik	Portal	n	Pro (%)	Non-Pro (%)	Netral (%)
Kriminalitas dan Penegakan Hukum	Detik.com	155	34,2	1,3	64,5
Kriminalitas dan Penegakan Hukum	Kompas.com	1.009	24,9	9,7	65,4
Politik dan Pemerintahan	Detik.com	55	36,4	1,8	61,8
Politik dan Pemerintahan	Kompas.com	761	28,6	10,8	60,6
Ekonomi dan Bisnis	Detik.com	1	100,0	0,0	0,0
Ekonomi dan Bisnis	Kompas.com	721	29,3	9,0	61,7
Olahraga dan Hiburan	Detik.com	1	100,0	0,0	0,0
Olahraga dan Hiburan	Kompas.com	656	3,5	1,4	95,1
Korupsi dan Penegakan Hukum	Detik.com	74	25,7	8,1	66,2
Korupsi dan Penegakan Hukum	Kompas.com	327	17,4	11,6	70,9

Secara keseluruhan, mayoritas artikel pada kedua portal cenderung diberi label Tidak Relevan/Netral (69,3%). Proporsi label Pro-Pemerintah lebih tinggi pada Detik.com (32,9%) dibandingkan Kompas.com (21,9%), sementara proporsi label Non-Pro-Pemerintah justru lebih tinggi pada Kompas.com (8,4%) dibandingkan Detik.com (3,1%). Pada tataran topik, kecenderungan Pro-Pemerintah paling menonjol pada topik Politik dan Pemerintahan di kedua portal (36,4% pada Detik.com; 28,6% pada Kompas.com), sedangkan topik Olahraga dan Hiburan pada Kompas.com hampir seluruhnya bersifat netral (95,1%), konsisten dengan sifat topik tersebut yang minim keterkaitan langsung dengan kebijakan pemerintah. Kecenderungan ini merefleksikan pola pembedaan pada periode data yang dianalisis (1–31 Mei 2026) dan tidak dimaksudkan sebagai penilaian permanen atas afiliasi institusi media. Interpretasi tetap perlu ditakar

secara proporsional dengan keterbatasan penelitian, yakni periode pengamatan yang terbatas, pemangkasan teras berita, potensi bias model, serta ukuran sampel acuan evaluasi yang masih terbatas ($n = 12$) sebagaimana diuraikan pada Sub-bab 3.2.

Dominasi label netral (69,3%) dapat dipahami karena sebagian besar pemberitaan harian melaporkan peristiwa secara faktual tanpa mengambil posisi eksplisit terhadap pemerintah; pola serupa—ketika kelas netral menyerap proporsi terbesar—juga dijumpai pada studi deteksi sikap atas pemberitaan yang terpolarisasi [14]. Asimetri yang teramati, yaitu proporsi Pro-Pemerintah yang lebih tinggi pada Detik.com dan proporsi Non-Pro-Pemerintah yang lebih tinggi pada Kompas.com, konsisten dengan gagasan *framing* bahwa pilihan penonjolan dan sudut pandang dapat mencerminkan kecenderungan redaksi yang berbeda [1]. Karena keluaran model bertumpu pada teras berita dan bukan teks penuh, sebagian sinyal sikap yang termuat pada bagian akhir artikel berpeluang tidak tertangkap; hal ini, bersama ukuran subset Detik.com yang relatif kecil ($n = 286$), menuntut kehati-hatian dalam menggeneralisasi temuan. Secara praktis, hasil ini memperlihatkan bahwa pendekatan berbasis LLM tanpa pelabelan manual dapat menjadi instrumen pemantauan wacana media yang skalabel, selama interpretasinya tetap memperhatikan keterbatasan metodologis di atas.

4. KESIMPULAN

Penelitian ini telah menerapkan dan mengevaluasi sebuah *pipeline* analisis pemberitaan daring berbasis model bahasa besar DeepSeek untuk menemukan isu dominan (*topic discovery*) dan mengklasifikasikan keberpihakan portal berita terhadap pemerintah (*stance detection*) tanpa memerlukan pelabelan manual maupun penetapan kata kunci. Tujuan penelitian terjawab: dari 27.789 artikel yang dihimpun, pendekatan yang memadukan ekstraksi teras berita (*inverted pyramid*) dan *two-pass prompting* yang didukung *context caching* terbukti mampu menangani korpus berskala masif secara efisien, dengan biaya pemrosesan seluruh korpus hanya sebesar USD 1,56 menggunakan model deepseek-v4-flash atau diperkirakan sekitar 82% lebih hemat dibanding pemrosesan teks utuh. Penemuan topik menghasilkan lima isu utama, yaitu Kriminalitas dan Penegakan Hukum, Politik dan Pemerintahan, Ekonomi dan Bisnis, Olahraga dan Hiburan, serta Korupsi dan Penegakan Hukum. Pemetaan keberpihakan memperlihatkan bahwa pada periode 1–31 Mei 2026, pemberitaan Detik.com cenderung lebih Pro-Pemerintah (32,9%), sedangkan Kompas.com menunjukkan proporsi Non-Pro-Pemerintah yang lebih tinggi (8,4%); adapun klasifikasi sikap mencapai akurasi 91,67% dan F1-Score makro 0,930 pada evaluasi terhadap *gold standard* hasil anotasi manual ($n = 12$). Secara keseluruhan, kombinasi DeepSeek dengan strategi *inverted pyramid prompting* dan pemrosesan dua lintasan menjadi solusi yang efektif dan hemat biaya untuk analisis wacana media berskala besar, sekaligus mengatasi keterbatasan metode pembelajaran mesin klasik dalam menangani topik tak dikenal dan sikap tersirat. Penelitian lanjutan disarankan untuk menguji generalisasi pendekatan ini pada model bahasa besar sumber terbuka (*open-source*) lain guna membandingkan kinerja dan biaya, memperluas cakupan data ke platform media sosial yang gaya bahasanya lebih beragam, serta memperhalus klasifikasi sikap melalui skema kelas yang lebih granular dan data acuan berukuran lebih besar disertai pengujian kesepakatan antar-anotator untuk menekan potensi bias.

REFERENCES

- [1] R. M. Entman, "Framing: Toward clarification of a fractured paradigm," *Journal of Communication*, vol. 43, no. 4, pp. 51–58, 1993, doi: 10.1111/j.1460-2466.1993.tb01304.x.
- [2] S. M. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, "SemEval-2016 Task 6: Detecting stance in tweets," in *Proc. 10th Int. Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, CA, USA, 2016, pp. 31–41, doi: 10.18653/v1/S16-1003.
- [3] W. Yin, J. Hay, and D. Roth, "Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach," in *Proc. 2019 Conf. Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, 2019, pp. 3914–3923, doi: 10.18653/v1/D19-1404.
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Jan. 2003. [Online]. Available: <https://jmlr.org/papers/v3/blei03a.html>
- [5] D. Küçük and F. Can, "Stance detection: A survey," *ACM Computing Surveys*, vol. 53, no. 1, pp. 1–37, Feb. 2020, doi: 10.1145/3369026.
- [6] T. B. Brown et al., "Language models are few-shot learners," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 1877–1901, arXiv: 2005.14165.
- [7] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, 2022, pp. 24824–24837, arXiv: 2201.11903.
- [8] DeepSeek-AI, "DeepSeek API documentation," DeepSeek, 2026. [Online]. Available: <https://api-docs.deepseek.com> [Accessed: Jun. 15, 2026].

- [9] H. Pöttker, "News and its communicative quality: The inverted pyramid—when and why did it appear?" *Journalism Studies*, vol. 4, no. 4, pp. 501–511, 2003, doi: 10.1080/1461670032000136596.
- [10] DeepSeek-AI, "DeepSeek-V3 technical report," arXiv:2412.19437, 2024, doi: 10.48550/arXiv.2412.19437.
- [11] F. Gilardi, M. Alizadeh, and M. Kubli, "ChatGPT outperforms crowd workers for text-annotation tasks," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 120, no. 30, p. e2305016120, 2023, doi: 10.1073/pnas.2305016120.
- [12] B. Wilie et al., "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," in *Proc. 1st Conf. Asia-Pacific Chapter Assoc. Comput. Linguistics and 10th Int. Joint Conf. Natural Language Processing (ACL-IJCNLP)*, Suzhou, China, 2020, pp. 843–857, doi: 10.18653/v1/2020.aacl-main.85.
- [13] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," arXiv:2203.05794, 2022, doi: 10.48550/arXiv.2203.05794.
- [14] M. Mets, A. Karjus, I. Ibrus, and M. Schich, "Automated stance detection in difficult topics and small languages: The challenging case of immigration in polarizing news media," *PLoS ONE*, vol. 19, no. 4, p. e0302380, 2024, doi: 10.1371/journal.pone.0302380.
- [15] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, 2022, pp. 22199–22213, arXiv: 2205.11916.