

Disetujui oleh:	
Ahmad Ridha S.Kom., M.S.	Dr. Ir. Sri Wahjuni M.T.



**KOLOKIU**  
**PROGRAM STUDI SARJANA ILMU KOMPUTER**  
**DEPARTEMEN ILMU KOMPUTER**  
**INSTITUT PERTANIAN BOGOR**

---

Nama	: Salma Nadhira Danuningrat
NIM	: G6401211065
Program Studi	: Sarjana Ilmu Komputer
Judul Penelitian	: Analisis <i>Framing</i> Media Berita Indonesia terhadap Genosida Palestina menggunakan <i>Latent Dirichlet Allocation</i> (LDA) dan <i>Deep Learning</i>
Pembimbing	: 1. Ahmad Ridha S.Kom., M.S. 2. Dr. Ir. Sri Wahjuni M.T.
Hari / Tanggal	: Kamis / 6 Februari 2025
Waktu	: 09.00 s/d 09.30 WIB
Tempat	: Online
Pembahas	: 1. Tan, Maria Putri Ariyani: G6401211049 2. Naufal Akbar Rahardjo: G6401211119 3. Halida Fiadnin: G6401211142

---

# ANALISIS *FRAMING* MEDIA BERITA INDONESIA TERHADAP GENOSIDA PALESTINA MENGGUNAKAN *LATENT DIRICHLET ALLOCATION* (LDA) DAN *DEEP LEARNING*

Salma Nadhira Danuningrat<sup>1</sup>, Ahmad Ridha<sup>2</sup>, Sri Wahjuni<sup>3</sup>

<sup>1</sup>Mahasiswa Sarjana Program Studi Ilmu Komputer, IPB

<sup>2</sup>Pembimbing 1, staf pengajar Departemen Ilmu Komputer IPB

<sup>3</sup>Pembimbing 2, staf pengajar Departemen Ilmu Komputer IPB

## Abstrak

SALMA NADHIRA DANUNINGRAT. Analisis *Framing* Media Berita Indonesia Terhadap Genosida Palestina Menggunakan *Latent Dirichlet Allocation* (LDA) dan *Deep Learning*. Dibimbing oleh Ahmad Ridha dan Sri Wahjuni.

*Framing* media dalam pemberitaan konflik Palestina-Israel berperan penting dalam membentuk persepsi publik dan meminimalkan bias. Analisis *framing* secara otomatis dengan *Natural Language Processing* (NLP) semakin berkembang karena efektivitasnya dalam memanfaatkan *neural network* berbasis *transformer*. Penelitian ini bertujuan untuk menganalisis *framing* berita terkait genosida Palestina menggunakan dataset terbaru yang berisi judul berita dari beberapa media online Indonesia, termasuk Kompas, Tempo, Detik, Republika, CNN Indonesia, dan CNBC Indonesia. Pemodelan topik dilakukan dengan *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi topik atau *frame* utama dalam kumpulan data berita. Hasil pemodelan ini kemudian digunakan sebagai dasar pelabelan manual *framing* berita, yang selanjutnya dimanfaatkan untuk membandingkan kinerja model *deep learning* LSTM, GRU, dan IndoBERT dalam mengklasifikasikan judul berita berdasarkan hasil *topic modeling* dengan LDA. Hasil penelitian ini diharapkan dapat memberikan wawasan mengenai pola bias dan narasi yang dibentuk oleh media berita di Indonesia, memperkaya literatur tentang *framing* media dalam konflik internasional, serta menjadi dasar bagi pengembangan model NLP yang lebih akurat dan relevan untuk bahasa Indonesia.

**Kata Kunci:** *Framing* Media, Genosida Palestina, *Latent Dirichlet Allocation*, *Deep Learning*, IndoBERT, *Natural Language Processing*

## PENDAHULUAN

### Latar Belakang

Realitas yang disajikan atau ditangkap melalui media massa tidak sepenuhnya mencerminkan realitas sebagaimana adanya, melainkan merupakan hasil konstruksi dari media tersebut (Siregar dan Qurniawati 2021). *Framing* sebuah cerita merupakan mekanisme media menyoroti dan memberikan penekanan pada aspek-aspek tertentu dari sebuah narasi sambil mengesampingkan aspek lainnya, yang pada akhirnya membentuk persepsi publik terhadap suatu konflik (Entman 1993).

Berbagai penelitian tentang *framing* media terkait konflik Israel-Palestina telah dilakukan. Siswanti (2019) membandingkan pemberitaan dua media di Indonesia, yakni CNN Indonesia dan Kompas. CNN Indonesia cenderung merefleksikan ideologi Amerika Serikat yang mendukung Israel, terlihat dari pemberitaannya yang tampak netral. Sebaliknya, Kompas lebih menunjukkan dukungannya terhadap Palestina, selaras dengan kebijakan luar negeri Indonesia yang secara konsisten membela Palestina. Dalam penelitian Ramadani (*et al.* 2024), analisis pemberitaan dari berbagai media *online* populer di Indonesia terkait pengeboman Israel terhadap Palestina pada 1–3 November 2023 menggunakan *framing* yang bervariasi. Berita dari Detik, CNN Indonesia, dan Viva cenderung berfokus pada korban dan kerusakan sehingga membentuk *framing* yang menggambarkan kekejaman Israel. Sementara itu, berita dari Kompas, Liputan6, dan CNBC menggunakan kata-kata yang menciptakan kesan negatif terhadap kelompok militan Hamas dan bahwa Hamas bertanggung jawab atas eskalasi konflik dan memicu serangan balasan dari Israel.

Analisis *framing* secara manual adalah tugas yang menantang dan memakan waktu, mengingat banyaknya jumlah media serta tingginya tingkat kompleksitas yang sering kali membutuhkan pemahaman mendalam terhadap nuansa dalam pemberitaan (Liu *et al.* 2019). Pendekatan non-teknis, seperti literasi media atau analisis politik adalah langkah efektif untuk menghadapi tantangan bias berita, tetapi memerlukan banyak waktu dan usaha (Hamborg 2023). Maka dari itu, *media framing analysis* secara otomatis, melalui teknik *Natural Language Processing* (NLP) mulai dikembangkan. Langkah pertamanya adalah penggunaan metode *unsupervised* seperti *topic modelling* untuk mengidentifikasi tema umum dalam kumpulan dokumen yang dapat merepresentasikan *frame*. Salah satu metode *topic modelling* adalah *Latent Dirichlet Allocation* (LDA), yang merupakan model generatif untuk menjelaskan sekumpulan observasi berdasarkan kelompok-kelompok tersembunyi. Kelompok-kelompok tersebut membantu menerangkan kemiripan yang ditemukan antara beberapa bagian data (Jing 2014). Selanjutnya, *frame* tersebut dapat digunakan secara langsung atau dijadikan acuan untuk pelabelan manual. *Frame* yang sudah diberi label ini kemudian diterapkan metode *supervised* untuk mengidentifikasi *frame* secara otomatis pada data yang belum berlabel (Nicholls 2020).

Secara umum, metode yang digunakan untuk mendeteksi *framing*, seperti *topic modelling*, *clustering*, analisis sentimen, atau kombinasi dari metode tersebut mampu menghasilkan wawasan yang relevan dan tematik. Namun,

hasilnya sering kali terlalu umum sehingga diragukan apakah metode tersebut benar-benar dapat disebut sebagai deteksi *framing* karena belum dapat menangkap nuansa semantik yang diuraikan oleh Entman (Jumle 2024). Pendekatan yang dapat mengatasi keterbatasan tersebut adalah model *neural network*, terutama model *transformer* seperti BERT, yang terbukti unggul dari metode tradisional seperti *Support Vector Machine* (SVM) dalam deteksi *frame* otomatis (Khanehzar et al. 2019). Melalui metode *deep learning* dengan *neural network*, identifikasi *frame* dari berbagai isu berita telah diterapkan. Verbytska (2024) menggunakan LDA bersamaan dengan BERT untuk menganalisis *frame* dari peliputan media barat dan timur terhadap perang Rusia-Ukraina. Hasil penelitiannya menunjukkan 18 topik dalam media timur dan 11 topik dalam media barat yang disimpulkan menjadi tujuh *frame* utama. Liu (et al. 2019) membandingkan *neural network* berbasis LSTM, GRU, dan BERT dalam mengidentifikasi *framing* berita Amerika Serikat terhadap topik kekerasan senjata. Model BERT berhasil mencapai akurasi *5-fold cross-validation* tertinggi sebesar 84,23% dalam mengidentifikasi sembilan jenis *frame* yang ditentukan berdasarkan literatur yang tersedia.

### Perumusan Masalah

*Framing* pemberitaan media secara manual sering kali memakan waktu yang signifikan dan memerlukan pemahaman mendalam terkait topik yang dibahas. Oleh karena itu, rumusan masalah dari penelitian ini adalah bagaimana *framing* peliputan genosida Palestina dalam media Indonesia dianalisis dengan pendekatan LDA dan *Deep Learning*?

### Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan dataset baru berisi judul-judul berita dari berbagai sumber berita Indonesia, yang dirancang untuk mendukung penelitian ini sekaligus dapat digunakan oleh publik.
2. Mengevaluasi efektivitas metode *Latent Dirichlet Allocation* (LDA) dalam melakukan topic modeling pada dataset tersebut.
3. Membandingkan kinerja beberapa model *deep learning* dalam mengklasifikasikan judul berita berdasarkan hasil topic modeling dengan LDA.

### Manfaat

Penelitian ini diharapkan dapat berkontribusi pada pengembangan NLP berbahasa Indonesia yang masih terbatas. Selain itu, penelitian ini juga diharapkan bermanfaat dalam peningkatan kesadaran dan pemahaman lebih mendalam tentang narasi dan bias media yang terdapat dalam berita Indonesia. Hasil penelitian ini juga dapat mendorong transparansi media dengan mengidentifikasi pola pemberitaan yang tidak berimbang sehingga dapat menjadi dasar untuk praktik jurnalistik yang lebih bertanggung jawab di Indonesia.

## Ruang Lingkup

Ruang lingkup dari penelitian ini adalah sebagai berikut:

1. Judul berita yang digunakan dalam dataset adalah berita publik yang dipublikasi dari 7 Oktober 2023 hingga 19 Januari 2025.
2. Dalam proses pelabelan sampel, setiap judul berita hanya diperbolehkan memiliki 1 *frame*.
3. Anotasi data manual dilakukan secara mandiri.

## METODE

### Data Penelitian

Judul merupakan elemen yang paling menonjol dalam memicu konsep-konsep tertentu yang secara semantik terkait di benak pembaca (Pan dan Kosicki 1993). Data yang digunakan dalam penelitian ini merupakan judul berita berbagai bentuk media (video, foto, infografik, dan artikel tertulis) yang diberi tagar #Palestina atau #Israel dan dipublikasi dalam rentang 471 hari, yaitu dari serangan kelompok militer Hamas pada 7 Oktober 2023 hingga kesepakatan tanggal permulaan gencatan senjata tanggal 19 Januari 2025. Judul berita diambil dari beberapa sumber media online yaitu Kompas, Tempo, Detik, Republika, CNN Indonesia, dan CNBC Indonesia. Pemilihan sumber berita tersebut bermaksud untuk membandingkan media dengan beragam nilai, segmen, dan lisensi. Setelah melalui proses pembersihan data, seluruh judul berita yang terkumpul akan diterapkan *topic modelling* menggunakan metode *Latent Dirichlet Allocation* (LDA). Proses ini bertujuan mengidentifikasi kecenderungan kata-kata yang muncul pada setiap topik untuk dijadikan acuan saat pelabelan *frame* sebagian judul berita secara manual.

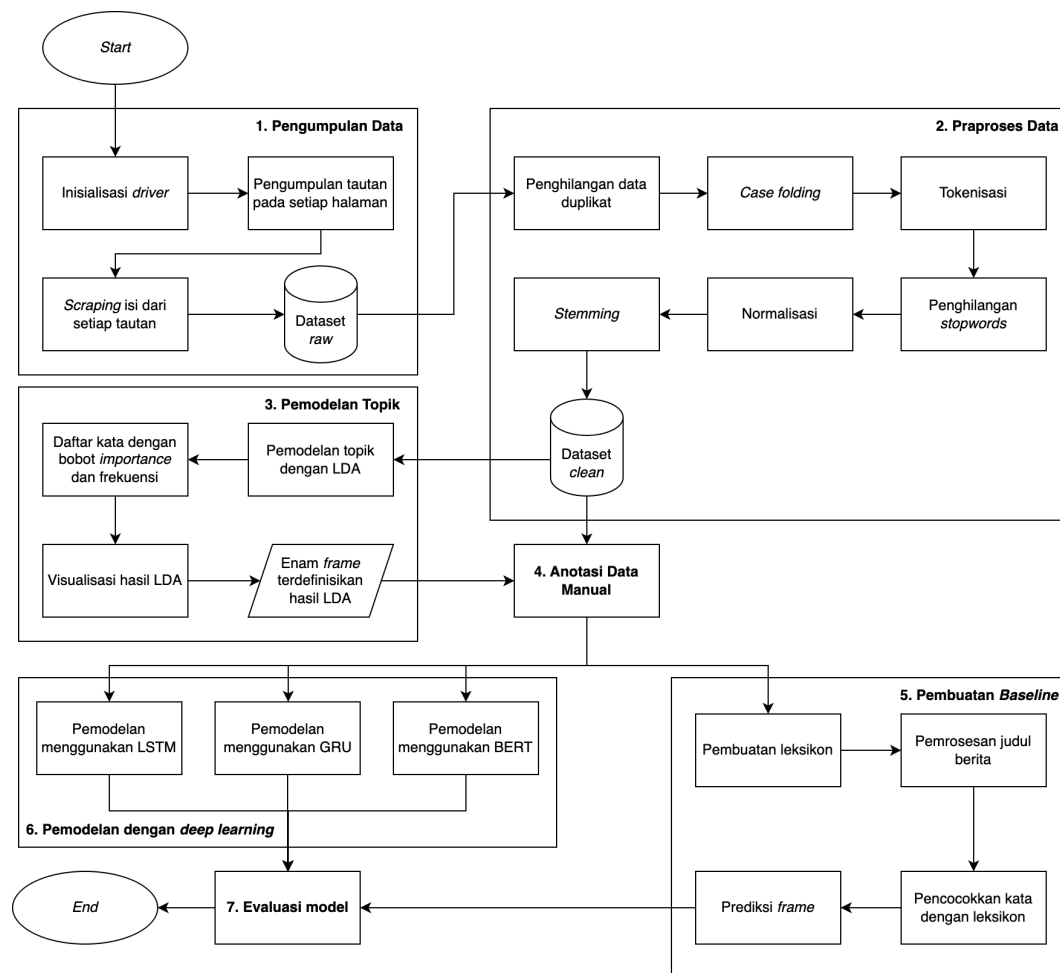
### Peralatan Penelitian

Penelitian ini menggunakan perangkat keras dan perangkat lunak sebagai berikut:

1. Perangkat keras berupa komputer dengan spesifikasi:
  - Intel CORE i5 8th Gen
  - RAM 8 GB, SSD 512 GB
2. Perangkat lunak yang digunakan antara lain:
  - Sistem Operasi macOS Monterey Versi 12.7.6
  - Bahasa pemrograman Python 3.9
  - Visual Studio Code 1.96.0
  - Google Colab
3. Library Python yang digunakan antara lain:
  - Pengumpulan data: BeautifulSoup4, Selenium WebDriver, SeleniumBase
  - Pemodelan topik dengan LDA: Pandas, Numpy, Swifter, Sastrawi, NLTK, Gensim, pyLDAvis

## Tahapan Penelitian

Penelitian ini dilaksanakan dalam tujuh tahapan yaitu pengumpulan data, praproses data, pemodelan topik, anotasi data manual, pembuatan *baseline*, pemodelan dengan *deep learning*, dan evaluasi model.



Gambar 1 Tahapan penelitian

### Pengumpulan Data

Data didapatkan dengan proses *scraping* keenam situs web berita publik yang memiliki tagar #Palestina atau #Israel. *Scraping* data menggunakan Python dengan library BeautifulSoup4 sebagai pengurai HTML, Selenium WebDriver untuk mengotomatiskan interaksi dalam web, serta SeleniumBase sebagai *framework* tambahan dari Selenium WebDriver dengan fitur tambahan untuk memudahkan proses otomasi web. Proses *scraping* data melalui empat tahapan, yakni inisialisasi *driver*, mengumpulkan tautan yang berada dalam setiap halaman web, *scraping* isi dari setiap tautan yang dikumpulkan, serta penyimpanan hasil *scraping* ke dalam file CSV. Hasil dari *scraping* data adalah satu *file* CSV yang berisi seluruh judul data yang terkumpulkan beserta atributnya yaitu: id, judul\_berita, tanggal\_publicasi, jenis\_media, sumber\_berita, dan penulis\_berita.

### **Praproses Data**

Prapemrosesan data sebelum pelatihan model membantu memastikan bahwa model bekerja secara adil dan transparan (Zelaya 2019). Praproses data juga merupakan langkah awal sebelum pengimplementasian LDA agar hasil *topic modelling* seakurat dan seoptimal mungkin. Praproses data menggunakan bahasa pemrograman Python dengan memanfaatkan beberapa *library*. Tahapan praproses meliputi penghilangan data duplikat; *case folding*, yaitu mengubah seluruh huruf menjadi huruf kecil menggunakan *library* Pandas; tokenisasi, yaitu penghapusan karakter tidak penting seperti angka, tanda baca, dan spasi, serta memecah teks menjadi bagian yang lebih kecil yang disebut "token" menggunakan *library* NLTK; pembersihan teks dari kata-kata yang termasuk dalam daftar *stopwords*, yang diperoleh dari modul *nlTK.corpus* dengan *stopwords* bahasa Indonesia yang telah tersedia; normalisasi, yaitu mengganti kata-kata tertentu dengan kata-kata lain yang memiliki makna serupa; serta *stemming* menggunakan *library* Sastrawi untuk mengubah kata menjadi bentuk dasarnya dengan menghilangkan awalan, akhiran, atau imbuhan lainnya.

### **Pemodelan Topik**

Setelah data terkumpul, *topic modelling* atau pemodelan topik diterapkan dengan metode LDA sebagai langkah *unsupervised* untuk menentukan topik-topik yang terkandung dalam data. Dalam penelitian ini, jumlah topik atau *frame* ditentukan terlebih dahulu yaitu sebanyak enam *frame* berdasarkan penelitian oleh Shahzad *et al.* (2023). Dalam penelitiannya, diidentifikasi enam topik dalam liputan konflik Israel-Palestina dalam RT News, Al-Jazeera, BBC, and CNN, yaitu *frame* "Conflict", "Human Interest", "Economic Consequence", "Mortality", "Responsibility", dan "Nationalization".

Setelah praproses data, LDA diterapkan menggunakan *library* Gensim. Hasil akhir dari pemodelan ini mencakup daftar seluruh kata dalam dataset beserta topik yang diidentifikasi, bobot *importance* yang merepresentasikan relevansi kata terhadap topiknya, serta frekuensi kemunculannya dalam data. Berdasarkan kumpulan kata dalam setiap topik, masing-masing topik diberikan definisi konkret yang kemudian digunakan sebagai "judul" topik tersebut dan menjadi acuan dalam proses anotasi data manual. Analisis lebih lanjut dilakukan dengan *library* pyLDavis guna memvisualisasikan distribusi jarak antar topik serta kata-kata dengan frekuensi tertinggi pada setiap topik.

### **Anotasi Data Manual**

Data dianotasi secara mandiri dengan mencocokkan setiap judul berita dalam sampel dengan salah satu dari keenam *frame* yang telah ditentukan saat pemodelan topik. Penentuan topik untuk setiap data dilakukan berdasarkan nilai bobot *importance tertinggi* dari kata-kata yang terdapat dalam judul berita. Judul berita kemudian dikategorikan ke dalam topik yang memiliki bobot relevansi tertinggi.

### **Pembuatan Baseline**

Penelitian ini menggunakan *baseline* berbasis metode leksikon yang dikembangkan oleh Field *et al.* (2018) untuk mendeteksi *frame* dalam berita. Metode ini dirancang untuk mengidentifikasi *frame* dalam teks dengan

mengandalkan hubungan antara kata-kata tertentu dan *frame* tertentu dalam sebuah korpus. Pendekatan ini terdiri dari beberapa langkah utama yang melibatkan pembuatan leksikon, pemrosesan judul berita, pencocokan kata dengan leksikon, dan evaluasi prediksi *frame*.

Proses dimulai dengan membangun leksikon *frame* untuk setiap *frame*  $F$ . Leksikon ini dibuat berdasarkan perhitungan *Pointwise Mutual Information* (PMI), yang digunakan untuk menentukan seberapa erat hubungan antara suatu kata  $w$  dengan *frame*  $F$ .

$$PMI(F, w) = \log \frac{P(F, w)}{P(F) \cdot P(w)} \quad (1)$$

Setelah nilai PMI dihitung untuk setiap kata dalam korpus, kata-kata dengan nilai PMI tertinggi dipilih untuk membentuk leksikon masing-masing *frame*. Field *et al.* (2018) menetapkan bahwa setiap leksikon terdiri dari 250 kata dengan nilai PMI tertinggi yang paling berkaitan dengan *frame* tersebut.

Setelah leksikon selesai dibuat, *frame* utama (*primary frame*) dari sebuah judul berita ditentukan dengan mencocokkan kata-kata dalam judul dengan kata-kata yang terdapat dalam leksikon setiap *frame*. Sebuah *frame* dianggap sebagai kandidat jika setidaknya tiga kata dalam judul tersebut ditemukan dalam leksikon *frame* tersebut. Jika lebih dari satu *frame* memenuhi syarat ini, maka *frame* dengan jumlah kata terbanyak yang ditemukan dalam leksikonnya dipilih sebagai *frame* utama.

### Pemodelan dengan *Deep Learning*

Pemodelan beserta parameternya mengacu pada penelitian Liu (*et al.* 2019) yang menggunakan tiga model deep learning untuk mendeteksi *frame* dalam judul berita, yaitu *Long Short-Term Memory* (LSTM), *Gated Recurrent Unit* (GRU), dan BERT. Penelitian ini mengalami modifikasi yaitu penggunaan IndoBERT, yaitu model BERT yang dilatih menggunakan korpus Bahasa Indonesia. Setiap model dilatih menggunakan dataset yang telah dianotasi secara manual dengan pembagian data *5-fold cross validation*.

Model LSTM dan GRU yang digunakan memiliki arsitektur *bidirectional* untuk menangkap konteks dari kedua arah teks. Mekanisme *attention* ditambahkan untuk memberikan bobot lebih pada kata-kata penting dalam menentukan *frame*. Representasi kata pada model akan menggunakan embedding GloVe berukuran 100 dimensi sebagai representasi awal kata-kata dalam teks. Parameter model mencakup 128 unit pada lapisan RNN, satu lapisan mekanisme *attention*, ukuran batch sebesar 128, dan jumlah langkah pelatihan sebanyak 2000 langkah. Proses pelatihan menggunakan optimizer Adam dengan laju pembelajaran sebesar 0,01. Model akan dilatih selama *5-fold cross validation*, dan hasil prediksi dibandingkan dengan label anotasi manual.

IndoBERT adalah model *transformer* berbasis BERT yang telah dilatih sebelumnya pada korpus besar berbahasa Indonesia. Parameter model akan mencakup maksimum panjang urutan 128 token, ukuran batch sebesar empat, dan laju pembelajaran  $2 \times 10^{-5}$ . Model dilatih selama sepuluh *epoch* dengan pendekatan *end-to-end*, termasuk pembaruan bobot pada semua parameter.





## DAFTAR PUSTAKA

- Entman RM. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*. 43(4):51–58. doi: 10.1111/j.1460-2466.1993.tb01304.x.
- Field A, Kliger D, Wintner S, Pan J, Jurafsky D, Tsvetkov Y. 2018. Framing and Agenda-setting in Russian News: A Computational Analysis of Intricate Political Strategies. Di dalam: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. doi:https://doi.org/10.18653/v1/d18-1393.
- Hamborg F. 2023. Revealing Media Bias in News Articles. *Springer Nature*. https://link.springer.com/book/10.1007/978-3-031-17693-7.
- Jing Q. 2014. Searching for economic effects of user specified events based on topic modelling and event reference [disertasi]. Wolfville: Acadia University.
- Jumle V, Makhortykh M, Sydorova M, Vziatysheva V. 2024. Finding frames with BERT: A transformer-based approach to generic news frame detection. ArXiv (Cornell University). doi: 10.48550/arxiv.2409.00272.
- Khanehzar S, Turpin A, Mikolajczak G. 2019. Modeling political framing across policy issues and contexts. Di dalam: The 17th Annual Workshop of the Australasian Language Technology Association. pp. 61–66.
- Liu S, Guo L, Mays K, Betke M, Wijaya DT. 2019. Detecting Frames in News Headlines and Its Application to Analyzing News Framing Trends Surrounding U.S. Gun Violence. Di dalam: Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL). pp. 504–514. doi: 10.18653/v1/k19-1047.
- Nicholls T, Culpepper PD. 2020. Computational Identification of Media Frames: Strengths, Weaknesses, and Opportunities. *Political Communication*. 1–23. doi: 10.1080/10584609.2020.1812777.
- Pan Z, Kosicki GM. 1993. Framing analysis: An approach to news discourse. *Political Communication*. 10(1):55–75.
- Ramadani MS, Kurniawan K, Fuadin A. 2024. Menguak Bias Media dalam Pemberitaan Konflik Israel-Palestina: Sebuah Analisis Konten Kritis. *Jurnal Onoma: Pendidikan, Bahasa, dan Sastra*. 10(1):887–905.
- Shahzad F, Qazi TA, Shehzad R. 2023. Framing of Israel and Palestine Conflict in RT news, Al-Jazeera, CNN & BBC News. *Global Digital & Print Media Review*. 6(2):1–14. doi:https://doi.org/10.31703/gdpmr.2023(vi-ii).01.
- Siregar AK, Qurniawati EF. 2022. Analisis Framing Pemberitaan Buzzer di tempo.co. *Journal of New Media and Communication*. 1(1):1–15. doi: 10.55985/jnmc.v1i1.1.
- Siswanti N. 2019. Analisis Framing Media: Studi Komparatif Media Online “CNN” dan “Kompas” Terkait Fenomena Kemanusiaan di Al-Aqsa Periode 20-23 Juli 2017. *Jurnal Riset Komunikasi*. 2(2):110–125.
- Verbytska A. 2024. Topic Modelling as a Method for Framing Analysis of News Coverage of the Russia-Ukraine War in 2022–2023. *Language & Communication*. 99:174–193. doi: 10.1016/j.langcom.2024.10.004.
- Zelaya C. 2019. Towards Explaining the Effects of Data Preprocessing on Machine Learning. Di dalam: 2019 IEEE 35th International Conference on Data Engineering (ICDE). 2086–2090. doi: 10.1109/icde.2019.00245.