

ANALISIS *FRAME* DAN KLASIFIKASI BERITA INDONESIA TERKAIT GENOSIDA PALESTINA DENGAN *LATENT* *DIRICHLET ALLOCATION* DAN *DEEP LEARNING*

SALMA NADHIRA DANUNINGRAT



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Analisis *Frame* dan Klasifikasi Judul Berita Media Indonesia terkait Genosida Palestina menggunakan *Latent Dirichlet Allocation* dan *Deep Learning*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2025

Salma Nadhira Danuningrat
G6401211065

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

ABSTRAK

SALMA NADHIRA DANUNINGRAT. Analisis *Frame* dan Klasifikasi Judul Berita Media Indonesia terkait Genosida Palestina menggunakan *Latent Dirichlet Allocation* dan *Deep Learning*. Dibimbing oleh AHMAD RIDHA dan SRI WAHJUNI.

Identifikasi *frame* yang terkandung dalam pemberitaan konflik Palestina-Israel oleh media Indonesia memiliki peran penting dalam membentuk persepsi publik dan menyaring bias informasi. Penelitian ini bertujuan untuk menganalisis *frame* dan mengklasifikasikan judul berita dari enam media online Indonesia dengan pendekatan *Natural Language Processing* (NLP). Hasil pemodelan topik dengan *Latent Dirichlet Allocation* (LDA) digunakan untuk melabeli data, yang selanjutnya diklasifikasi dengan model *deep learning* LSTM, Bi-LSTM, GRU, Bi-GRU, dan IndoBERT. Hasil penelitian menunjukkan bahwa dari delapan label klasifikasi yang dikembangkan dari enam topik hasil LDA, IndoBERT memiliki kinerja terbaik dengan nilai F1 rata-rata makro tertimbang sebesar 81,20% sementara model RNN mengalami *overfitting* pada data berlabel otomatis hasil LDA. Penelitian ini menegaskan potensi metode berbasis *Transformer* dalam memahami konteks pada data serta relevansi pendekatan NLP dalam merepresentasi isu dalam pemberitaan media.

Kata kunci: *deep learning*, genosida palestina, indobert, *latent dirichlet allocation*, *natural language processing*, pemodelan topik

ABSTRACT

SALMA NADHIRA DANUNINGRAT. Frame Analysis and Classification of Indonesian Media News Titles on the Palestinian Genocide using Latent Dirichlet Allocation and Deep Learning. Supervised by AHMAD RIDHA and SRI WAHJUNI.

The identification of *frames* in the coverage of the Palestine-Israel conflict by Indonesian media plays an important role in shaping public perception. This study aims to analyze *frames* and classify news titles from six Indonesian online media outlets using Natural Language Processing (NLP). Data labeled from the results of topic modeling with Latent Dirichlet Allocation (LDA) are then classified using LSTM, Bi-LSTM, GRU, Bi-GRU, and IndoBERT. The results of this study show that of the eight classification labels developed from the six topics provided by the LDA results, IndoBERT performs the best with a weighted macro-averaged F1-score of 81.20% while the RNN model experiences overfitting on automatically labeled data from LDA. This study confirms the potential of Transformer-based methods in understanding context in data and the relevance of NLP in representing issues in media reporting.

Keywords: deep learning, indobert, latent dirichlet allocation, natural language processing, palestinian genocide, topic modeling

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

ANALISIS *FRAME* DAN KLASIFIKASI BERITA INDONESIA TERKAIT GENOSIDA PALESTINA DENGAN *LATENT DIRICHLET ALLOCATION* DAN *DEEP LEARNING*

SALMA NADHIRA DANUNINGRAT

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Skripsi : Analisis *Frame* dan Klasifikasi Judul Berita Media Indonesia terkait Genosida Palestina menggunakan *Latent Dirichlet Allocation* dan *Deep Learning*

Nama : Salma Nadhira Danuningrat
NIM : G6401211065

Disetujui oleh

Pembimbing 1:
Ahmad Ridha S.Kom., M.S.

Pembimbing 2:
Dr. Ir. Sri Wahjuni M.T.

Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:
Dr. Sony Hartono Wijaya, S.Kom., M.Kom.
19810809 200812 1 002

Tanggal Ujian:
15 Juli 2025

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan November 2024 sampai bulan Juni 2025 ini ialah *Natural Language Processing* (NLP), dengan judul “Analisis *Frame* dan Klasifikasi Judul Berita Media Indonesia terkait Genosida Palestina menggunakan *Latent Dirichlet Allocation* dan *Deep Learning*”.

Terima kasih penulis ucapkan pada:

- a. Bapak Ahmad Ridha, S.Kom., M.S. dan Dr. Ir. Sri Wahjuni, M.T. selaku dosen pembimbing, atas bimbingan, dukungan, serta kepercayaan yang telah diberikan kepada penulis dan topik penelitian ini. Terima kasih atas setiap saran, masukan, dan diskusi yang sangat berarti selama proses penyusunan skripsi ini. Ucapan terima kasih juga penulis sampaikan kepada dosen penguji, Ibu Lailan Sahrina Hasibuan S.Kom., M.Kom., atas saran dan koreksi demi kebaikan tugas akhir ini.
- b. Kedua orang tua tercinta, Ibu Natasha Kania Idris dan Bapak Budi Darmawan Danuningrat, atas doa yang tak henti-hentinya, dukungan tanpa syarat, serta kasih sayang yang tiada tara. Semoga Allah senantiasa melimpahkan kesehatan dan keberkahan umur kepada Mamah dan Papah.
- c. Rekan-rekan Ilmu Komputer 58, MaximIAAS, teman-teman MSIB Batch 7 dan NSQM Telkomsel, Digistar Intern Telkom, serta teman-teman seperjuangan yang telah mewarnai hari-hari kuliah penulis, khususnya Cheisha, Maria, Bella, dan Tata, terima kasih atas kebersamaan dan semangatnya.
- d. Sahabat-sahabat tercinta: Bilqis, Aimee, Haya, Alia, Moja, Shasa, Zanira, serta sahabat lainnya yang tidak bisa disebutkan satu per satu, terima kasih atas dukungan yang tak pernah surut, serta kebersamaannya di setiap suka dan duka.
- e. Narrel, Disha, Nella, serta Niken, Aki, dan seluruh keluarga yang telah menjadi bagian tak terpisahkan dalam kehidupan penulis.
- f. Terakhir, kepada sahabat dan nenek tercinta, Almarhumah Herna Pruintina Danuningrat, yang selalu menaruh harapan besar kepada penulis, yang dengan penuh kasih senantiasa mengantarkan penulis dari satu jenjang pendidikan ke jenjang berikutnya, semoga penulis dapat menjadi salah satu jalur amal jariah bagi beliau.

Sebagai penutup, penulis memanjatkan doa bagi rakyat Palestina yang hingga kini terus mengalami penindasan. Semoga Allah subhanahu wa ta'ala senantiasa melimpahkan rahmat, kekuatan, dan perlindungan kepada mereka. Semoga kita yang jauh dari sana tidak pernah berhenti berpihak pada keadilan, dan menjadikan teknologi sebagai jalan untuk menyuarakan kebenaran. Semoga karya ilmiah ini dapat memberikan manfaat bagi pihak yang membutuhkan serta berkontribusi bagi pengembangan ilmu pengetahuan.

Bogor, Juli 2025

Salma Nadhira Danuningrat

DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
II METODE	5
2.1 Tahapan Penelitian	5
2.2 Pengumpulan Data	6
2.3 Praproses Data untuk LDA	6
2.4 Pemodelan Topik dengan LDA	8
2.5 Anotasi Data Manual	10
2.6 Praproses Data untuk Model <i>Deep Learning</i>	10
2.7 Pemodelan dengan <i>Deep Learning</i>	10
2.8 Evaluasi Model <i>Deep Learning</i>	13
III HASIL DAN PEMBAHASAN	14
3.1 Pengumpulan Data	14
3.2 Praproses Data untuk LDA	14
3.2.1 <i>Case folding</i> dan penghapusan tanda baca	15
3.2.2 Normalisasi	15
3.2.3 Tokenisasi	15
3.2.4 Penghilangan <i>stopwords</i>	16
3.2.5 Deteksi <i>bigram</i> dan <i>trigram</i>	16
3.2.6 <i>Stemming</i>	17
3.2.7 <i>Term Frequency–Inverse Document Frequency</i> (TF-IDF)	17
3.3 Pemodelan Topik dengan LDA	18
3.4 Anotasi Data Manual	21
3.5 Praproses Data untuk Model <i>Deep Learning</i>	24
3.6 Pemodelan dengan <i>Deep Learning</i>	24
3.7 Evaluasi Model <i>Deep Learning</i>	25
IV SIMPULAN DAN SARAN	29
4.1 Simpulan	29
4.2 Saran	29
DAFTAR PUSTAKA	30
RIWAYAT HIDUP	41

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1 Sebaran data berdasarkan sumber berita	14
2 Jumlah data berdasarkan batas minimal jumlah kata	14
3 Hasil <i>case folding</i> pada contoh data	15
4 Hasil normalisasi pada contoh data	15
5 Hasil tokenisasi pada contoh data	16
6 Hasil penghilangan <i>stopwords</i> pada contoh data	16
7 Hasil deteksi <i>bigram</i> dan <i>trigram</i> pada contoh data	17
8 Sampel dari hasil deteksi <i>bigram</i> dan <i>trigram</i>	17
9 Hasil <i>stemming</i> pada contoh data	17
10 Hasil deskripsi <i>frame</i> secara heuristik	21
11 Contoh judul berita yang dihilangkan dengan alasannya	22
12 Jumlah judul berita per <i>frame</i> per sumber berita	23
13 Hasil akurasi dan loss dengan <i>stratified 5-fold cross validation</i>	25
14 Hasil <i>F1-score</i> setiap <i>frame</i> setiap model	26
15 Hasil akurasi <i>stratified 5-fold cross validation</i> ketiga <i>dataset</i>	28

DAFTAR GAMBAR

1 Tahapan penelitian	6
2 Tahapan praproses data untuk LDA	6
3 Representasi LDA oleh Blei <i>et al.</i> (2003)	9
4 Arsitektur LSTM (Zhang <i>et al.</i> 2023)	11
5 Arsitektur GRU (Zhang <i>et al.</i> 2023)	11
6 Arsitektur model <i>Transformer</i> (Vaswani <i>et al.</i> 2017)	12
7 Grafik jumlah topik berdasarkan nilai <i>coherence</i>	18
8 Visualisasi 5, 7, dan 6 topik	19
9 Hasil LDA dengan 8 topik	19
10 Hasil pencarian alpha dan eta berdasarkan nilai <i>coherence</i>	20
11 <i>Word cloud</i> untuk 6 topik berdasarkan kata dengan <i>importance</i> tertinggi	21
12 Diagram lingkaran sebaran frame setiap sumber berita	24
13 Hasil <i>confusion matrix</i> dengan BERT per subset data	26
14 Hasil akurasi ketiga dataset pada LSTM dan GRU	27

DAFTAR LAMPIRAN

1 Daftar kata normalisasi untuk praproses data sebelum LDA dan pemodelan dengan <i>deep learning</i>	36
2 Daftar <i>stopwords</i> tambahan untuk dihilangkan saat praproses dikarenakan berfrekuensi tinggi atau bersifat <i>Named Entity</i>	38

3 Hasil LDA dengan enam topik berupa sebaran kata, *importance*, dan *word count* untuk masing-masing topik

39

@Hak cipta milik IPB University

IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.