

SKRIPSI
TOPIC MODELING PADA JURNAL PINTER
MENGGUNAKAN METODE *LATENT DIRICHLET*
ALLOCATION



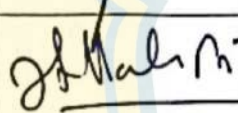
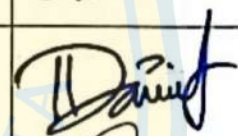
Disusun oleh:
DAVIN ARAZI
1512622031

PROGRAM STUDI
PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER
FAKULTAS TEKNIK
2026

LEMBAR PENGESAHAN TUGAS AKHIR
BENTUK : SKRIPSI

Nama Mahasiswa : Davin Arazi
NIM : 1512622031
Program Studi/Fakultas : Pendidikan Teknik Informatika dan Komputer / Teknik
Judul Penelitian : Topic Modeling Pada Jurnal Pinter Menggunakan Metode Latent Dirichlet Allocation

ini telah dipertahankan dihadapan dewan penguji dan dinyatakan **Lulus / Tidak Lulus**)*
Ujian Tugas Akhir oleh Tim Penguji pada Tanggal 9 Juni 2026
Tim Penguji,

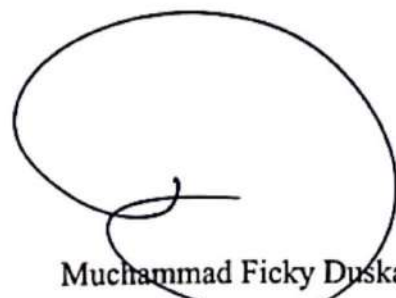
Ketua Penguji	Muchammad Ficky Duskarnaen, M.Sc NIDN. 0024097304	
Anggota Penguji (Penguji Ahli 1)	Hamidillah Ajie, S.Si., M.T NIDN. 0024087402	
Anggota Penguji (Penguji Ahli 2)	Diat Nurhidayat, S.Pd., M.T.I NIDN. 0019088303	
Anggota Penguji (Pembimbing 1)	Dr. Widodo, S.Kom., M.Kom. NIDN. 0025037206	
Anggota Penguji (Pembimbing 2)	Neng Ayu Herawati, S.Pd., M.T. NUPTK.1252776677230153	

Mengetahui,
Dekan Fakultas Teknik

Koordinator Program Studi
Pendidikan Teknik Informatika dan
Komputer



Prof. Dr. Nengeng Siti Silfi Ambarwati, S.Si., Apt., M.Si.
NIDN. 0029027204



Muchammad Ficky Duskarnaen, M.Sc
NIDN. 0024097304

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini merupakan karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 2 Juni 2026
Yang membuat pernyataan



Davin Arazi
No. Reg. 1512622031



KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
PERPUSTAKAAN DAN KEARSIPAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Davin Arazi
NIM : 1512622031
Fakultas/Prodi : Fakultas Teknik – Pendidikan Teknik Informatika dan Komputer
Alamat email : davinarazi07@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan dan Kearsipan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (... ..)

yang berjudul :

TOPIC MODELLING PADA JURNAL PINTER MENGGUNAKAN METODE LATENT DIRICHLET

ALLOCATION

Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan dan Kearsipan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 7 Agustus 2026

Penulis

Davin Arazi

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi berjudul "*Topic Modeling* pada Jurnal PINTER Menggunakan Metode Latent Dirichlet Allocation". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Pendidikan pada Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta.

Penulis menyadari bahwa penyusunan skripsi ini tidak terlepas dari doa, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Ayah dan Ibu Penulis, atas kasih sayang, doa, dan dukungan yang tidak pernah berhenti menjadi kekuatan bagi penulis dalam menyelesaikan setiap proses perkuliahan.
2. Bapak Muhammad Ficky Duskarnaen, M.Sc., selaku Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer, atas dukungan dan arahan selama masa studi.
3. Bapak Dr. Widodo, S.Kom., M.Kom., selaku Dosen Pembimbing I, atas ilmu, arahan, dan masukan yang sangat berharga selama proses penelitian dan penyusunan skripsi ini.
4. Ibu Neng Ayu Herawati, S.Pd., M.T., selaku Dosen Pembimbing II, atas arahan, bantuan, serta dukungan dan semangat yang senantiasa diberikan dalam mendampingi penulis selama proses penyusunan skripsi ini.
5. Naufal, Shafira, Risma, Enjel dan Jasmin, sahabat sejak masa SMA yang senantiasa memberikan semangat, kebersamaan, dan keceriaan.
6. Farhan, Indah, Azizah, serta rekan-rekan PTIK lainnya yang telah memberikan dukungan, kebersamaan, dan semangat selama masa studi.
7. Pihak-pihak lain yang turut membantu dalam penyelesaian skripsi ini, baik secara langsung maupun tidak langsung, yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi penyempurnaan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan.

Jakarta, 2 Juni 2026



Davin Arazi
1512622031



ABSTRAK

Davin Arazi, *Topic Modeling* pada Jurnal PINTER Menggunakan Metode *Latent Dirichlet Allocation*. Dosen Pembimbing: Dr. Widodo, S.Kom., M.Kom. dan Neng Ayu Herawati, S.Pd., M.T. Program Studi Pendidikan Teknik Informatika dan Komputer. Fakultas Teknik. Universitas Negeri Jakarta. 2026.

Meningkatnya volume publikasi ilmiah dalam Jurnal PINTER menyebabkan kesulitan dalam mengidentifikasi distribusi tema penelitian secara manual, sementara belum terdapat mekanisme otomatis berbasis komputasi yang diadopsi untuk pemetaan topik pada jurnal tersebut. Kondisi ini mendorong penerapan pendekatan *text mining* berbasis *topic modeling* menggunakan metode *Latent Dirichlet Allocation* (LDA) yang mampu mengidentifikasi struktur tematik tersembunyi dalam korpus dokumen secara probabilistik, sekaligus membandingkan tiga metode representasi fitur yaitu TF-IDF, *FastText*, dan kombinasi TF-IDF+*FastText* untuk mengetahui pendekatan yang menghasilkan kualitas topik terbaik. Proses penelitian dilakukan melalui tahapan *Knowledge Discovery in Databases (KDD)* yang meliputi *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*, dengan menggunakan dataset 160 artikel *full text* dari Jurnal PINTER Volume 1 Nomor 1 hingga Volume 8 Nomor 2. Tahap *preprocessing* mencakup *data cleansing*, *case folding*, *tokenization*, *stemming*, dan *stopword removal* menggunakan pustaka Sastrawi. Pengujian dilakukan pada dua skenario data, yaitu berbasis volume dan berbasis jurnal, dengan variasi jumlah topik $K = 3$, $K = 4$, dan $K = 5$. Evaluasi model menggunakan *coherence score* (C_v) untuk mengukur keterkaitan semantik antar kata dalam topik dan *perplexity* untuk mengukur kemampuan generalisasi model. Hasil penelitian menunjukkan bahwa metode TF-IDF menghasilkan nilai *perplexity* tertinggi pada seluruh konfigurasi (rata-rata 233,990 berbasis volume), mengindikasikan keterbatasan representasi berbasis frekuensi dalam memodelkan distribusi probabilistik kata. Metode *FastText* menunjukkan performa yang lebih seimbang dengan titik optimal pada $K = 4$ berbasis volume (*coherence* = 0,7379; *perplexity* = 38,701). Kombinasi TF-IDF+*FastText* terbukti menghasilkan performa terbaik secara konsisten, dengan model optimal pada $K = 5$ berbasis volume (*coherence* = 0,6709; *perplexity* = 16,793) dan $K = 4$ berbasis jurnal (*coherence* = 0,6203; *perplexity* = 40,060). Berdasarkan topik yang terbentuk, tema riset dominan dalam Jurnal PINTER mencakup jaringan komputer dan analisis kualitas layanan, media pembelajaran berbasis teknologi, pengembangan sistem dan aplikasi, serta evaluasi dan pengukuran dalam Pendidikan teknik informatika.

Kata Kunci: *topic modeling*, *Latent Dirichlet Allocation*, TF-IDF, *FastText*, *coherence score*, *perplexity*, Jurnal PINTER

ABSTRACT

Davin Arazi, *Topic Modeling* on PINTER Journal Using Latent Dirichlet Allocation Method. Supervisors: Dr. Widodo, S.Kom., M.Kom. and Neng Ayu Herawati, S.Pd., M.T. Informatics and Computer Engineering Education Study Program. Faculty of Engineering. Universitas Negeri Jakarta. 2026.

The growing volume of scientific publications in PINTER Journal has made it increasingly difficult to manually identify the distribution of research themes, while no automated computational mechanism has been adopted for topic mapping in the journal. This condition prompted the application of a text mining approach based on *topic modeling* using the Latent Dirichlet Allocation (LDA) method, which is capable of probabilistically identifying hidden thematic structures within a document *corpus*, while simultaneously comparing three feature representation methods TF-IDF, *FastText*, and the TF-IDF+*FastText* combination to determine the approach that yields the best topic quality. The research was conducted through the Knowledge Discovery in Databases (KDD) stages comprising selection, preprocessing, transformation, data mining, and evaluation, using a dataset of 160 full-text articles from PINTER Journal Volume 1 Number 1 through Volume 8 Number 2. The preprocessing stage included data cleansing, case folding, tokenization, stemming, and stopword removal using the Sastrawi library. Experiments were conducted under two data scenarios volume-based and journal based with topic count variations of $K = 3$, $K = 4$, and $K = 5$. Model evaluation employed *coherence score* (C_v) to measure *semantic* relatedness among words within a topic, and *perplexity* to assess the model's generalization capability. Results show that TF-IDF produced the highest *perplexity* values across all configurations (average 233.990 in the volume-based scenario), indicating the limitations of frequency-based representation in modeling probabilistic word distributions. *FastText* demonstrated more balanced performance with an optimal point at $K = 4$ in the volume-based scenario (*coherence* = 0.7379; *perplexity* = 38.701). The TF-IDF+*FastText* combination consistently produced the best performance, with the optimal model at $K = 5$ in the volume-based scenario (*coherence* = 0.6709; *perplexity* = 16.793) and $K = 4$ in the journal-based scenario (*coherence* = 0.6203; *perplexity* = 40.060). Based on the identified topics, the dominant research themes in PINTER Journal include computer networks and quality of service analysis, technology-based learning media, system and application development, and evaluation and measurement in informatics and computer engineering education.

Keywords: *topic modeling*, Latent Dirichlet Allocation, TF-IDF, *FastText*, *coherence score*, *perplexity*, PINTER Journal

DAFTAR ISI

LEMBAR PENGESAHAN UJIAN SKRIPSI.....	i
LEMBAR PERNYATAAN	ii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	iii
KATA PENGANTAR.....	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Identifikasi Masalah	4
1.3 Pembatasan Masalah	5
1.4 Perumusan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
1.6.1 Manfaat Teoritis	5
1.6.2 Manfaat Praktis	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Kerangka Teoritik	7
2.1.1 Pengertian Jurnal Ilmiah	7
2.1.2 Jurnal PINTER.....	9
2.1.3 <i>Text Mining</i>	10
2.1.4 <i>Text Preprocessing</i>	11
2.1.5 <i>FastText</i>	13
2.1.6 Term Frequency – Inverse Document Frequency TF-IDF	14
2.1.7 <i>Topic Modeling</i>	16
2.1.8 <i>Latent Dirichlet Allocation (LDA)</i>	17
2.1.9 <i>Coherence Score</i>	20
2.1.10 <i>Perplexity</i>	21
2.1.11 <i>Doc2Bow</i>	22
2.1.12 <i>PyLDAvis</i>	23
2.1.13 <i>Knowledge Discovery in Databases (KDD)</i>	24

2.2 Penelitian Relevan.....	25
2.3 Kerangka Berpikir.....	36
BAB III METODOLOGI PENELITIAN	38
3.1 Tempat dan Waktu Penelitian.....	38
3.2 Alat dan Bahan Penelitian.....	38
3.2.1 Alat Penelitian.....	38
3.2.2 Bahan Penelitian.....	39
3.3 Langkah-langkah Penelitian.....	40
3.4 Prosedur Penelitian.....	43
3.4.1 <i>Selection</i>	43
3.4.2 <i>Preprocesssing</i>	44
3.4.3 <i>Transformation</i>	46
3.4.4 <i>Data Mining</i>	47
3.4.5 Interpretation / Evaluation.....	49
3.5 Teknik Pengumpulan Data.....	53
BAB IV HASIL DAN PEMBAHASAN.....	55
4.1 Hasil Pengumpulan Data.....	55
4.2 Tahap <i>Preprocessing</i> Data	57
4.3 Pembangunan Model <i>Topic Modeling</i> (LDA)	64
4.3.1 Membangun <i>dictionary</i> dan <i>corpus</i>	64
4.3.2 Mengkonversi TF-IDF	66
4.3.3 Mengkonversi <i>FastText</i>	67
4.3.4 Menggabungkan TF-IDF dan <i>FastText</i>	69
4.4 Hasil <i>Topic Modeling</i>	70
4.4.1 Hasil Skenario 1 (TF-IDF).....	77
4.4.2 Hasil Skenario 2 (<i>FastText</i>)	84
4.4.3 Hasil Skenario 3 (TF-IDF + <i>FastText</i>).....	92
4.4.4 Kesimpulan Hasil <i>Topic Modeling</i>	101
4.5 Pembahasan.....	103
4.6 Ringkasan Hasil	105
BAB V KESIMPULAN DAN SARAN	107
5.1 Kesimpulan	107
5.2 Saran.....	108
DAFTAR PUSTAKA	109
LAMPIRAN.....	113

DAFTAR GAMBAR

Gambar 2. 1 Kerangka Berpikir	37
Gambar 3. 1 Diagram Alir Penelitian	42
Gambar 4. 1 Grafik <i>Coherence Score</i> Per Volume.....	73
Gambar 4. 2 Grafik <i>Coherence Score</i> Per Jurnal	74
Gambar 4. 3 Grafik <i>Perplexity</i> Terbaik Per Volume	75
Gambar 4. 4 Grafik <i>Perplexity</i> Terbaik per Jurnal.....	76
Gambar 4. 5 Hasil <i>Topic Modeling</i> dari TF-IDF Per Volume	79
Gambar 4. 6 Hasil <i>Topic Modeling</i> dari TF-IDF Per Jurnal	80
Gambar 4. 7 <i>PyLDAvis</i> TF-IDF Per Volume	80
Gambar 4. 8 <i>Wordcloud</i> TF-IDF per Volume.....	81
Gambar 4. 9 <i>PyLDAvis</i> TF-IDF Per Jurnal	82
Gambar 4. 10 <i>Wordcloud</i> TF-IDF per Jurnal.....	83
Gambar 4. 11 Hasil <i>Topic Modeling</i> dari FastText per Volume	86
Gambar 4. 12 Hasil <i>Topic Modeling</i> dari <i>FastText</i> Per Jurnal.....	87
Gambar 4. 13 <i>PyLDAvis FastText</i> per Volume.....	87
Gambar 4. 14 <i>Wordcloud FastText</i> per Volume	88
Gambar 4. 15 <i>PyLDAvis FastText</i> per Jurnal.....	89
Gambar 4. 16 <i>Wordcloud FastText</i> per Jurnal	90
Gambar 4. 17 Hasil <i>Topic Modeling</i> dari TF-IDF + <i>FastText</i> per Volume	95
Gambar 4. 18 Hasil <i>Topic Modeling</i> dari TF-IDF + <i>FastText</i> per Jurnal	95
Gambar 4. 19 <i>PyLDAvis</i> TF-IDF + <i>FastText</i> per Volume	96
Gambar 4. 20 <i>Wordcloud</i> TF-IDF + <i>FastText</i> per Volume	97
Gambar 4. 21 <i>PyLDAvis</i> TF-IDF + <i>FastText</i> per Jurnal	98
Gambar 4. 22 <i>Wordcloud</i> TF-IDF + <i>FastText</i> per Jurnal	99

DAFTAR TABEL

Tabel 2. 1 Contoh Proses Data Cleansing	11
Tabel 2. 2 Contoh Proses Case Folding	12
Tabel 2. 3 Contoh Proses Tokenization	12
Tabel 2. 4 Contoh Proses Stemming	13
Tabel 2. 5 Contoh Proses Stopword Removal.....	13
Tabel 2. 6 Penelitian Relevan.....	30
Tabel 3. 1 Spesifikasi Macbook Air M2	38
Tabel 3. 2 Spesifikasi Software.....	38
Tabel 4. 1 Distribusi Data Penelitian	57
Tabel 4. 2 Hasil dari <i>Cleansing</i>	57
Tabel 4. 3 Hasil dari <i>Case Folding</i>	58
Tabel 4. 4 Hasil dari <i>Tokenization</i>	59
Tabel 4. 5 Hasil dari <i>Stemming</i>	60
Tabel 4. 6 Hasil dari <i>Stopwords Removal</i>	62
Tabel 4. 7 Hasil <i>Preprocessing</i> Dokumen	62
Tabel 4. 8 Contoh <i>Dictionary</i> (<i>Term</i> → ID)	65
Tabel 4. 9 Contoh <i>Corpus</i> (ID <i>Term</i> → Frekuensi)	65
Tabel 4. 10 Contoh TF (<i>Term Frequency</i>)	66
Tabel 4. 11 Contoh TF-IDF.....	67
Tabel 4. 12 Contoh Representasi <i>FastText</i>	68
Tabel 4. 13 <i>Vector FastText</i> Awal	69
Tabel 4. 14 Hasil Kombinasi TF-IDF dan <i>FastText</i>	69
Tabel 4. 15 Rekap <i>Hasil Topic Modeling</i> per Volume	71
Tabel 4. 16 Rekap <i>Hasil Topic Modeling</i> per Jurnal.....	71
Tabel 4. 17 Hasil Evaluasi TF-IDF berdasarkan Volume	78
Tabel 4. 18 Hasil Evaluasi TF-IDF berdasarkan Jurnal	79
Tabel 4. 19 Hasil Evaluasi <i>FastText</i> per Volume.....	85
Tabel 4. 20 Hasil Evaluasi <i>FastText</i> per Jurnal.....	86
Tabel 4. 21 Hasil Evaluasi TF-IDF + <i>FastText</i> per Volume.....	93
Tabel 4. 22 Hasil Evaluasi TF-IDF + <i>FastText</i> per Jurnal	94
Tabel 4. 23 Perbandingan Nilai Evaluasi Seluruh Skenario	101

DAFTAR LAMPIRAN

Lampiran 1. Lembar Pernyataan Kelayakan Judul/Tema Skripsi.....	113
Lampiran 2. Surat Tugas	114
Lampiran 3. Lembar Konsultasi Skripsi.....	115
Lampiran 4. Kartu Kehadiran Seminar Proposal Skripsi	119
Lampiran 5. Biodata Peserta Ujian Skripsi	120
Lampiran 6. Surat Permohonan Ujian Skripsi.....	123
Lampiran 7. Surat Persetujuan Dosen Pembimbing Skripsi	124
Lampiran 8. Surat Persetujuan Dosen Pembimbing Akademik.....	126
Lampiran 9. Pra-Transkrip Akademik.....	127
Lampiran 10. Surat Pernyataan Menyelesaikan Mata Kuliah Wajib	130
Lampiran 11. Dataset Penelitian	131
Lampiran 12. Hasil Penelitian.....	132

