

Topik Modelling Skripsi Prodi Teknik Lingkungan di Jawa Timur Menggunakan Metode LDA

Nurul Khoirinnisa^a, Khalid^b, Noor Wahyudi^c, Yunita Ardilla^d

^{a,b,c,d}Program Studi Sistem Informasi, Universitas Islam Negeri Sunan Ampel Surabaya

^aniskhoirin05@gmail.com, ^bkhalid@uinsa.ac.id, ^cn.wahyudi@uinsa.ac.id, ^dyunita.ardilla@uinsa.ac.id

Submitted: 02-10-2025, Reviewed: 06-10-2025, Accepted 25-09-2025

<https://doi.org/10.47233/jteksis.v7i4.2278>

Abstract

In higher education, the undergraduate thesis represents a tangible contribution of students to the advancement of scientific knowledge. In the field of Environmental Engineering, student research often focuses on issues relevant to both local and global environmental challenges. This study aims to analyze the main themes of undergraduate theses from Environmental Engineering students in East Java using the Latent Dirichlet Allocation (LDA) method. The dataset consists of 644 thesis titles and abstracts collected from three universities: UIN Sunan Ampel Surabaya (UINSA), Institut Teknologi Sepuluh Nopember (ITS), and Universitas Nahdlatul Ulama Sidoarjo (UNUSIDA). Through the application of LDA, several dominant topics were identified, including waste management, water quality, air pollution reduction, and the implementation of environmental treatment technologies. The resulting topic model achieved a coherence score of 0.4885, indicating a reasonably good level of topic consistency and interpretability. The analysis demonstrates that LDA effectively reveals research patterns and clusters key issues of interest among students. These findings not only provide an overview of student research trends but also serve as a basis for curriculum development, research planning, and academic decision-making. Therefore, this study contributes to improving educational quality and mapping the research direction in the field of Environmental Engineering.

Keywords: LDA; Environmental Engineering; Student Thesis; Topic Analysis; Research Trends.

Abstrak

Dalam dunia pendidikan tinggi, skripsi merupakan salah satu bentuk nyata kontribusi mahasiswa terhadap pengembangan ilmu pengetahuan. Pada bidang Teknik Lingkungan, penelitian mahasiswa seringkali berfokus pada isu-isu yang relevan dengan tantangan lingkungan baik di tingkat lokal maupun global. Penelitian ini bertujuan untuk menganalisis tema utama dalam skripsi mahasiswa Program Studi Teknik Lingkungan di Jawa Timur dengan menggunakan metode *Latent Dirichlet Allocation* (LDA). Data penelitian diperoleh dari 644 judul dan abstrak skripsi yang dikumpulkan dari tiga universitas, yaitu UIN Sunan Ampel Surabaya (UINSA), Institut Teknologi Sepuluh Nopember (ITS), dan Universitas Nahdlatul Ulama Sidoarjo (UNUSIDA). Melalui penerapan LDA, sejumlah topik dominan berhasil diidentifikasi, di antaranya pengelolaan limbah, kualitas air, pengurangan pencemaran udara, serta penerapan teknologi pengolahan lingkungan. Model topik yang dihasilkan memiliki nilai coherence score sebesar 0,4885, yang menunjukkan kualitas pemodelan yang cukup baik dalam menggambarkan keterkaitan antar topik. Hasil analisis menunjukkan bahwa LDA mampu mengungkap pola penelitian mahasiswa dan mengelompokkan isu-isu yang menjadi perhatian utama. Temuan ini tidak hanya memberikan gambaran tren penelitian mahasiswa, tetapi juga dapat dijadikan dasar dalam perumusan strategi pengembangan kurikulum, perencanaan penelitian, serta pengambilan keputusan akademik. Dengan demikian, penelitian ini berkontribusi pada upaya peningkatan kualitas pendidikan dan pemetaan arah penelitian di bidang Teknik Lingkungan.

Keywords: LDA; Teknik Lingkungan; Skripsi Mahasiswa; Analisis Topik; Trend Penelitian.

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Program studi Teknik Lingkungan memiliki peran penting dalam menghasilkan solusi inovatif terhadap permasalahan lingkungan yang semakin kompleks[1]. Dalam proses pendidikan, skripsi menjadi salah satu wujud nyata kontribusi mahasiswa dalam pengembangan ilmu pengetahuan, terutama di bidang Teknik Lingkungan. Di Jawa timur memiliki tantangan lingkungan yang beragam, topik-topik skripsi di Prodi teknik Lingkungan menunjukkan variasi yang mencerminkan kebutuhan lokal, mulai dari pengelolaan limbah hingga mitigasi dampak

perubahan iklim[2][3]. Namun, analisis yang sistematis terhadap trend topik skripsi masih terbatas, sehingga menyulitkan dalam menentukan arah riset yang relevan dan prioritas keilmuan yang harus dikembangkan.

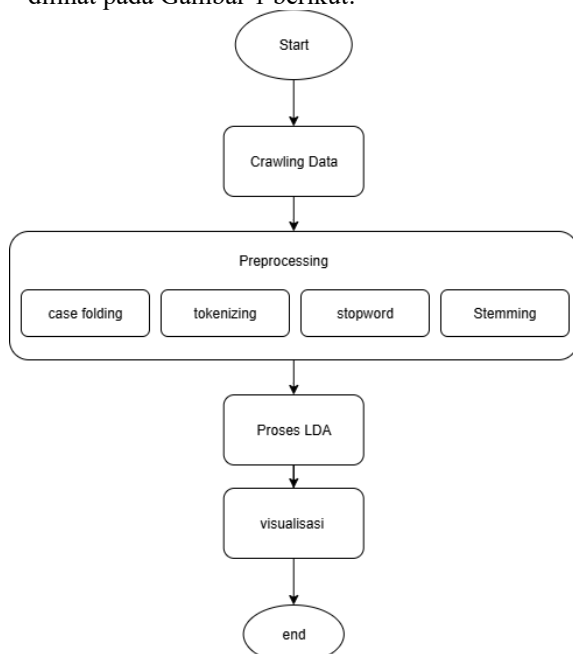
Salah satu pendekatan yang dapat digunakan untuk menggali pola dan trend dari sekumpulan dokumen teks adalah Latent Dirichlet Allocation (LDA). Metode ini memungkinkan identifikasi topik utama secara otomatis dari sekumpulan data teks yang besar, sehingga sangat cocok untuk menganalisis trend dalam karya tulis ilmiah seperti

skripsi[4][5]. Penelitian ini bertujuan untuk menerapkan metode LDA dalam menganalisis topik-topik skripsi mahasiswa Teknik Lingkungan di Jawa Timur[6].

Hasil analisis diharapkan mampu memberikan wawasan baru mengenai perkembangan riset di bidang ini, mengidentifikasi kebutuhan penelitian di masa depan, serta mendukung pengambilan keputusan akademik untuk meningkatkan relevansi dan kualitas kurikulum[5]. Selain itu, temuan penelitian ini dapat digunakan sebagai dasar untuk mengarahkan mahasiswa dan dosen dalam memilih topik riset yang lebih strategis dan berdampak signifikan bagi penyelesaian masalah lingkungan di Jawa. Dengan mengetahui topik-topik yang sering maupun jarang digunakan, Institusi dapat mengidentifikasi dan kesenjangan penelitian, serta merumuskan langkah-langkah strategis untuk mendorong eksplorasi topik yang kurang diminati tetapi memiliki potensi besar untuk dikembangkan[3].

METODE PENELITIAN

Metodologi penelitian adalah proses yang terdiri dari beberapa langkah yang disusun secara sistematis untuk memecahkan masalah penelitian dengan landasan ilmiah. Tahapan penelitian dapat dilihat pada Gambar 1 berikut:



Gambar 1. Metode Penelitian

2.1. Crawling data

Pada tahap ini, crawling data sudah dilakukan secara maksimal akan tetapi kendala-kendala atau hambatan yang ada membuat crawling data hanya dapat diambil dari 3 website digilib kampus, yaitu UINSA, ITS, dan UNUSIDA. Hambatan-hambatan

ini meliputi adanya akses terbatas, link tidak valid, dan batasan-batasan lainnya.

Pada tahapan crawling data didapatkan hasil dari tahun 2018 sampai 2024 untuk UINSA, 2019 sampai 2024 untuk ITS, dan 2022 sampai 2024 untuk UNUSIDA.

2.2 Pre-processing

Pada tahap Pre-processing data merubah bentuk dokumen kedalam bentuk yang memudahkan dan mempercepat proses dalam dokumen yang relevan. Setiap langkah pra-pemrosesan bertujuan untuk membuat indeks dari kumpulan dokumen. Proses pengindeksan dilakukan untuk membedakan satu dokumen dari dokumen lainnya. Pembuatan indeks perlu melibatkan konsep pemrosesan linguistik yang bertujuan untuk mengambil kata-kata penting dari setiap dokumen yang diwakili sebagai sekumpulan kata. Konsep pemrosesan linguistik mencakup case folding, remove symbol, stopword, stemming, tokenizing[7][8].

A. Case folding

Case folding adalah proses mengubah semua huruf dalam teks menjadi satu bentuk standar, biasanya huruf kecil (lowercase). Tujuannya agar “Komputer”, “komPuTer”, “KOMPUTER” dianggap sama saat pemrosesan teks, sehingga tidak terjadi duplikasi term akibat perbedaan kapitalisasi.

B. Remove symbol

Remove symbol digunakan untuk menghapus simbol-simbol pada teks. Seperti tanda tutup kurung dan buka kurung “()” dan lainnya.

C. Stopword

Stopword adalah kata-kata yang dianggap kurang memiliki makna semantik untuk tujuan tertentu dalam pemrosesan teks, misalnya kata penghubung (“dan”, “yang”, “di”, “ke” dalam bahasa Indonesia), kata depan, artikel, atau kata sangat umum yang sering muncul di banyak dokumen dan tidak membedakan kategori atau topik. Stopword removal atau filtering adalah proses penghapusan kata-kata tersebut agar teks yang diproses menjadi lebih “bersih” dan efisien.

D. Stemming

Stemming adalah proses mengubah kata-berimbuhan ke bentuk dasar (root / stem) dan menghilangkan afiks (awalan, akhiran, sisipan, konfiks) agar variasi kata yang berbeda bentuk tetapi memiliki makna dasar yang sama dianggap satu. Contohnya “bermain”, “memainkan”, “pemain” → menjadi

“main”. Proses ini membantu mengurangi variasi kata di dalam teks sehingga mempermudah analisis, indeks, atau klasifikasi[7].

E. Tokenizing

Tokenizing adalah tahap memotong / memecah teks menjadi unit-unit kecil (token), biasanya kata (word tokens), terkadang juga karakter atau frasa, tergantung aplikasinya. Pemotongan dilakukan berdasarkan pemisah seperti spasi, tanda baca, dan karakter pemisah lainnya. Contoh: dari kalimat “Saya pergi ke pasar” menjadi token “Saya”, “pergi”, “ke”, “pasar”.

2.3 Proses LDA

Proses pemodelan topik bertujuan untuk mendapatkan distribusi kata yang membentuk topik dan dokumen tertentu. Pemodelan topik terdiri dari dua langkah. Langkah pertama adalah melakukan pemodelan topik dengan menambah dan mengurangi jumlah topik. Langkah kedua adalah melakukan pemodelan topik berdasarkan jumlah literasi. Hasil dari kedua pemodelan topik kemudian dianalisis dengan membandingkan kata-kata di setiap kluster dalam topik dan melihat visualisasi dari pemodelan LDA tersebut. Proses pemodelan topik dapat dilakukan berulang kali selama rentang jumlah topik dan jumlah iterasi yang ditetapkan[9]. Latent Dirichlet Allocation (LDA) adalah salah satu metode dalam pemodelan topik. Model topik LDA adalah pembelajaran mesin tanpa pengawasan. Model ini bermanfaat untuk menemukan informasi yang tidak terlihat di dalam banyak dokumen. Proses ini dapat dilakukan dengan menggunakan python, setelah mengaktifkan paket “LdaModel” dalam pustaka gensim. Paket “LdaModel” digunakan untuk memprediksi kemungkinan kemunculan kata dalam dokumen[10][11]. Hasilnya berupa grafik yang menunjukkan topik dalam data yang diteliti.

2.4 Visualisasi

Tahap visualisasi ini dilakukan agar dapat dengan mudah dianalisis. Visualisasi menggunakan pyLDAvis dirancang untuk menafsirkan topik dalam model topik yang sesuai dengan kumpulan data teks[12].

HASIL DAN PEMBAHASAN

3.1. Pengumpulan data

Pengumpulan data pada penelitian ini dengan melakukan crawling data dari situs web digilib UINSA, ITS, dan UNUSIDA. Data yang diambil berupa abstraksi dari tugas akhir Teknik Lingkungan dari tahun 2018-2024. Akan tetapi

pada dilaksanakannya crawling data sering menemukan keterbatasan akses sehingga data yang didapat berupa tahun 2018 sampai 2024 untuk UINSA, 2019 sampai 2024 untuk ITS, dan 2022 sampai 2024 untuk UNUSIDA. Total data yang didapat dari pengumpulan ini mencapai 644 abstraksi. Dataset ini kemudian digunakan sebagai korpus untuk membuat pemodelan topik. Dataset terlebih dahulu dibersihkan dan kemudian diubah menjadi vektor TF-IDF. Vektor TF-IDF ini selanjutnya menjadi input untuk model LDA. Dalam pemodelan LDA, diperlukan parameter input yang berupa jumlah topik. Dalam studi ini, jumlah topik ditentukan dengan mencari nilai koherensi[13].

Tahun	Judul	Abstrak
2018	Studi tingkat partisipasi masyarakat Kelurahan...	Permasalahan sampah banyak terjadi di kota-kot...
2018	Perencanaan teknis Tempat Pengolahan Sampah (T...	Sampah menjadi salah satu permasalahan yang sa...
2018	Studi resapan air hujan melalui Lubang Resapan...	Kurangnya ruang terbuka hijau dapat mengakbat...
2018	Korelasi antara ruang terbuka hijau dengan kon...	Ruang Terbuka Hijau (RTH) merupakan area mema...
2019	Perencanaan Unit Pengolahan Lumpur Aluminium d...	Instalasi Pengolahan Air Minum X adalah Instal...
2019	Evaluasi Air Valve pada Jaringan Pipa Utama Di...	PDAM Kota Malang merupakan perusahaan penyedia...
2019	Kajian Mengenal Umr Tanaman Terhadap Reduksi ...	Konsentrasi CO2 yang meningkat dapat menyebabk...
2019	Alternatif Revitalisasi Instalasi Pengolahan L...	Pemerintah Kota Malang berkomitmen untuk menga...
2019	Kajian Pengendalian Partikel Flokulan dengan Hy...	Unit Hydrocyclone adalah alat yang digunakan u...
2019	Kombinasi Proses Pra-Komposting Dan Vermikomo...	Komposting konvensional yang berlangsung secar...
2019	Penentuan Status Mutu Air Sungai Kalimas Denga...	Sungai Kalimas merupakan salah satu anak Sunga...
2019	Evaluasi Kinerja Instalasi Pengolahan Air Limb...	Rumah sakit Universitas Tanjungpura Pontianak ...
2019	Kombinasi Proses Pra-Komposting dan Vermikomo...	Pengomposan material organik secara alami mem...
2019	Kajian Dampak Proses Produksi Minyak Bumi Terh...	Pemanasan global merupakan suatu fenomena glob...
2019	Analisis hubungan konsentrasi Total Suspended ...	Pencemar Udara Total Suspended Particulate (TSP...
2019	Studi Sistem Pengelolaan Sampah Kecamatan Sela...	Seiring dengan peningkatan kepadatan penduduk,...
2019	Pemanfaatan Limbah Abu Ketel (Ash Boiler) Untu...	Jumlah abu ketel yang dihasilkan PG. Kreboong...
2019	Kajian Teknis Dan Sosial Pengelolaan Air Limba...	Masyarakat di beberapa desa di Kecamatan Sedat...
2019	Perencanaan Pengelolaan Persampahan di Kampus VI...	UIN Sunan Ampel (UDINSA) merupakan salah satu p...
2019	Optimalisasi rute pengangkutan sampah TPA Lemp...	Timbulan sampah kabupaten Lumajang sebesar 203...
2019	Evaluasi kualitas udara karbon monoksida (co) ...	Menurut Damara dkk. (2017), gas CO yang masuk ...
2019	Upaya Minimisasi dan pengelolaan limbah Padat ...	RSU Haji Surabaya tengah melakukan upaya minis...
2019	Perencanaan sisten transmisi dan distribusi ai...	Kabupaten Hanggarai merupakan daerah dataran t...
2019	Evaluasi pengelolaan sampah di Rusunawa Guloma...	Keberadaan suatu Rusunawa kerap menimbulkan mas...
2024	Efisiensi aplikasi eco-enzyme terhadap air ter...	Terjadinya pencemaran dalam air berdampak terh...
2024	Analisis kadar PM10 di udara ambient serta kelu...	Terminal Purabaya merupakan terminal bus terbe...
2024	KAJIAN TIMBULAN SAMPAH DAN ANALISIS KAPASITAS ...	Tingkat pertumbuhan penduduk berpengaruh pada ...

Gambar 2. Sampel Data Penelitian

3.2. Preprocessing

Tahap yang bertujuan untuk mempersiapkan data sebelum dianalisis menggunakan LDA. Preprocessing dilakukan dengan menggunakan aplikasi Jupyter Notebook di Vs Code. Langkah pada proses ini dimulai dengan mengunduh dan mengaktifkan beberapa library yang dibutuhkan. Tahap pada pre-processing data dibagi menjadi lima tahap. Tahapan tersebut diantaranya case folding, remove symbols, tokenizing, stopwords removal, stemming.

Tahapan ini dimulai dengan melakukan case folding yaitu dengan mengubah setiap kata dengan bentuk standar atau huruf besar ke huruf kecil. Hasilnya seperti gambar 3.

STEP 1: Case Folding	Abstrak	Case_Folding
40	Berbagai kegiatan manusia akan menghasilkan li...	berbagai kegiatan manusia akan menghasilkan li...
34	Berdasarkan hasil uji pendahuluan yang telah d...	berdasarkan hasil uji pendahuluan yang telah d...
573	Rumah sakit atau institusi layanan kesehatan m...	rumah sakit atau institusi layanan kesehatan m...
48	Kota Yogyakarta adalah ibu kota dan pusat peme...	kota yogyakarta adalah ibu kota dan pusat peme...
218	Sepanjang daerah aliran sungai (das) kecamatan...	sepanjang daerah aliran sungai (das) kecamatan...
430	Lahan penimbunan TPA Ngipik terbagi menjadi du...	lahan penimbunan tpa ngipik terbagi menjadi du...
280	Keberadaan mikroplastik di perairan sedang men...	keberadaan mikroplastik di perairan sedang men...
458	At the Tambakrejo Material Recovery Facility, ...	at the tambakrejo material recovery facility, ...
370	Semakin tingginya permintaan pasar terhadap ik...	semakin tingginya permintaan pasar terhadap ik...
579	Aktivitas industri berhubungan dengan peningka...	aktivitas industri berhubungan dengan peningka...
476	Perkembangan ilmu pengetahuan dan teknologi da...	perkembangan ilmu pengetahuan dan teknologi da...
230	Pada tahun 2021 tercatat Indonesia diperkirakan...	pada tahun 2021 tercatat indonesia diperkirakan...
407	One issue that must be addressed, particularly...	one issue that must be addressed, particularly...
432	The growth of the population in Surabaya city...	the growth of the population in surabaya city...
13	Pemanasan global merupakan suatu fenomena glob...	pemanasan global merupakan suatu fenomena glob...
216	Desa Pagerwojo terletak di Kecamatan Buduran d...	desa pagerwojo terletak di kecamatan buduran d...
603	Menurut BPS Jawa Timur tahun 2023 Kabupaten Tu...	menurut bps jawa timur tahun 2023 kabupaten tu...
395	Pengolahan air limbah harus sesuai dengan peru...	pengolahan air limbah harus sesuai dengan peru...
11	Pada tahun 2021 tercatat Indonesia diperkirakan...	pada tahun 2021 tercatat indonesia diperkirakan...
179	Keberadaan Ruang Terbuka Hijau (RTH) di kota K...	keberadaan ruang terbuka hijau (rth) di kota k...

Gambar 3. Hasil Proses Case Folding

Tahapan selanjutnya yaitu remove symbol hal ini digunakan untuk menghapus simbol-simbol yang dapat mengganggu proses LDA. Hasilnya seperti gambar 4.

```

--- STEP 2: Remove Symbols ---
Case_Folding Remove_Symbols
40 berbagai kegiatan manusia akan menghasilkan li... berbagai kegiatan manusia akan menghasilkan li...
34 berdasarkan hasil uji pendahuluan yang telah d... berdasarkan hasil uji pendahuluan yang telah d...
573 rumah sakit atau institusi layanan kesehatan m... rumah sakit atau institusi layanan kesehatan m...
48 kota yogyakarta adalah ibu kota dan pusat peme... kota yogyakarta adalah ibu kota dan pusat peme...
218 sepanjang daerah aliran sungai (das) kecamatan... sepanjang daerah aliran sungai (das) kecamatan...
430 lahan penimbunan tpa ngipik terbagi menjadi du... lahan penimbunan tpa ngipik terbagi menjadi du...
280 keberadaan mikroplastik di perairan sedang men... keberadaan mikroplastik di perairan sedang men...
458 at the tambakrejo material recovery facility s... at the tambakrejo material recovery facility s...
370 semakin tingginya permintaan pasar terhadap ik... semakin tingginya permintaan pasar terhadap ik...
579 aktivitas industri berhubungan dengan peninka... aktivitas industri berhubungan dengan peninka...
476 perkembangan ilmu pengetahuan dan teknologi da... perkembangan ilmu pengetahuan dan teknologi da...
230 pada tahun 2021 tercatat indonesia diperkirakan... pada tahun 2021 tercatat indonesia diperkirakan...
407 one issue that must be addressed, particularly... one issue that must be addressed particularly...
432 the growth of the population in surabaya city... the growth of the population in surabaya city...
13 pemanasan global merupakan suatu fenomena glob... pemanasan global merupakan suatu fenomena glob...
216 desa pagerwojo terletak di kecamatan buduran d... desa pagerwojo terletak di kecamatan buduran d...
603 menurut bps jawa timur tahun 2023 kabupaten tu... menurut bps jawa timur tahun 2023 kabupaten tu...
385 pengolahan air limbah harus sesuai dengan peru... pengolahan air limbah harus sesuai dengan peru...
11 rumah sakit universitas tangjungpura pontianak... rumah sakit universitas tangjungpura pontianak...
179 keberadaan ruang terbuka hijau (rth) di kota k... keberadaan ruang terbuka hijau rth di kota ken...
% Generate + Code + Markdown

```

Gambar 4. Hasil Remove Symbol

Tahapan selanjutnya adalah tokenizing digunakan untuk memecah teks menjadi unit-unit kecil seperti kata. Hasilnya seperti gambar 5.

```

--- STEP 3: Tokenizing ---
Remove_Symbols Tokenizing
40 berbagai kegiatan manusia akan menghasilkan li... [berbagai, kegiatan, manusia, akan, menghasilkan...
34 berdasarkan hasil uji pendahuluan yang telah d... [berdasarkan, hasil, uji, pendahuluan, yang, t...
573 rumah sakit atau institusi layanan kesehatan m... [rumah, sakit, atau, institusi, layanan, keseh...
48 kota yogyakarta adalah ibu kota dan pusat peme... [kota, yogyakarta, adalah, ibu, kota, dan, pus...
218 sepanjang daerah aliran sungai das kecamatan t... [sepanjang, daerah, aliran, sungai, das, kecam...
430 lahan penimbunan tpa ngipik terbagi menjadi du... [lahan, penimbunan, tpa, ngipik, terbagi, menj...
280 keberadaan mikroplastik di perairan sedang men... [keberadaan, mikroplastik, di, perairan, sedana...
458 at the tambakrejo material recovery facility s... [at, the, tambakrejo, material, recovery, faci...
370 semakin tingginya permintaan pasar terhadap ik... [semakin, tingginya, permintaan, pasar, terhad...
579 aktivitas industri berhubungan dengan peninka... [aktivitas, industri, berhubungan, dengan, pen...
476 perkembangan ilmu pengetahuan dan teknologi da... [perkembangan, ilmu, pengetahuan, dan, teknolo...
230 pada tahun tercatat indonesia diperkirakan me... [pada, tahun, tercatat, indonesia, diperkirakan...
407 one issue that must be addressed particularly... [one, issue, that, must, be, addressed, partic...
432 the growth of the population in surabaya city... [the, growth, of, the, population, in, surabay...
13 pemanasan global merupakan suatu fenomena glob... [pemanasan, global, merupakan, suatu, fenomena...
216 desa pagerwojo terletak di kecamatan buduran d... [desa, pagerwojo, terletak, di, kecamatan, bud...
603 menurut bps jawa timur tahun kabupaten tuban... [menurut, bps, jawa, timur, tahun, kabupaten...
385 pengolahan air limbah harus sesuai dengan peru... [pengolahan, air, limbah, harus, sesuai, denga...
11 rumah sakit universitas tangjungpura pontianak... [rumah, sakit, universitas, tangjungpura, ponti...
179 keberadaan ruang terbuka hijau rth di kota ken... [keberadaan, ruang, terbuka, hijau, rth, di, k...

```

Gambar 5. Hasil Tokenizing

Tahapan selanjutnya adalah stopwords dengan menghapus kata-kata yang dianggap kurang bermakna seperti “dan”, “di”, “ke”, dan lainnya. Hasilnya seperti gambar 6.

```

--- STEP 4: Stopword Removal ---
Tokenizing Stopword_Remov
40 [berbagai, kegiatan, manusia, akan, menghasilk... [berbagai, kegiatan, manusia, akan, menghasilk...
34 [berdasarkan, hasil, uji, pendahuluan, dilakuk... [berdasarkan, hasil, uji, pendahuluan, dilakuk...
573 [rumah, sakit, atau, institusi, layanan, kesehat... [rumah, sakit, atau, institusi, layanan, kesehat...
48 [kota, yogyakarta, adalah, ibu, kota, dan, pus... [kota, yogyakarta, adalah, ibu, kota, pusat, pemerinta...
218 [sepanjang, daerah, aliran, sungai, das, kecam... [sepanjang, daerah, aliran, sungai, das, kecam...
430 [lahan, penimbunan, tpa, ngipik, terbagi, menj... [lahan, penimbunan, tpa, ngipik, terbagi, menj...
280 [keberadaan, mikroplastik, di, perairan, sedana... [keberadaan, mikroplastik, di, perairan, sedana...
458 [at, the, tambakrejo, material, recovery, faci... [at, the, tambakrejo, material, recovery, faci...
370 [semakin, tingginya, permintaan, pasar, terhad... [semakin, tingginya, permintaan, pasar, terhad...
579 [aktivitas, industri, berhubungan, dengan, pen... [aktivitas, industri, berhubungan, dengan, pen...
476 [perkembangan, ilmu, pengetahuan, dan, teknolo... [perkembangan, ilmu, pengetahuan, dan, teknolo...
230 [pada, tahun, tercatat, indonesia, diperkirakan... [pada, tahun, tercatat, indonesia, diperkirakan...
407 [one, issue, that, must, be, addressed, partic... [one, issue, that, must, be, addressed, partic...
432 [the, growth, of, the, population, in, surabay... [the, growth, of, the, population, in, surabay...
13 [pemanasan, global, merupakan, suatu, fenomena... [pemanasan, global, merupakan, suatu, fenomena...
216 [desa, pagerwojo, terletak, di, kecamatan, bud... [desa, pagerwojo, terletak, di, kecamatan, buduran...
603 [bps, jawa, timur, tahun, kabupaten, tuban, me... [bps, jawa, timur, tahun, kabupaten, tuban, me...
385 [pengolahan, air, limbah, harus, sesuai, peruntukanny... [pengolahan, air, limbah, harus, sesuai, peruntukanny...
11 [rumah, sakit, universitas, tangjungpura, ponti... [rumah, sakit, universitas, tangjungpura, ponti...
179 [keberadaan, ruang, terbuka, hijau, rth, kota... [keberadaan, ruang, terbuka, hijau, rth, kota...

```

Gambar 6. Hasil Stopword

Tahapan selanjutnya adalah stemming dengan menghapus kata berimbuhan menjadi kata dasar. Hasilnya seperti gambar 7.

```

--- STEP 5: Stemming ---
Stopword_Removal Stemming
40 [bagai, giat, manusia, hasil, limbah, buang, s... [bagai, giat, manusia, hasil, limbah, buang, s...
34 [dasar, hasil, uji, dahulu, laku, sampel, air,... [dasar, hasil, uji, dahulu, laku, sampel, air,...
573 [rumah, sakit, institusi, layan, sehat, rupa,... [rumah, sakit, institusi, layan, sehat, rupa,...
48 [kota, yogyakarta, ibu, kota, pusat, perintah... [kota, yogyakarta, ibu, kota, pusat, perintah...
218 [panjang, daerah, aliran, sungai, das, camat, ta... [panjang, daerah, aliran, sungai, das, camat, ta...
430 [lahan, timbun, tpa, ngipik, bagi, jadi, bagi,... [lahan, timbun, tpa, ngipik, bagi, jadi, bagi,...
280 [ada, mikroplastik, air, sedang, jadi, masalah... [ada, mikroplastik, air, sedang, jadi, masalah...
458 [at, the, tambakrejo, material, recovery, faci... [at, the, tambakrejo, material, recovery, faci...
370 [makin, tinggi, minta, pasar, ikan, air, tawar... [makin, tinggi, minta, pasar, ikan, air, tawar...
579 [aktivitas, industri, hubungan, tingkat, produks... [aktivitas, industri, hubungan, tingkat, produks...
476 [kembang, ilmu, tahu, teknologi, dorong, tingk... [kembang, ilmu, tahu, teknologi, dorong, tingk...
230 [tahun, catat, indonesia, kira, hasil, juta, t... [tahun, catat, indonesia, kira, hasil, juta, t...
407 [one, issue, that, must, be, addressed, partic... [one, issue, that, must, be, addressed, partic...
432 [the, growth, of, the, population, in, surabay... [the, growth, of, the, population, in, surabay...
13 [panasan, global, rupa, suatu, fenomena, globa... [panas, global, rupa, suatu, fenomena, globa...
216 [desa, pagerwojo, letak, camat, budur, kabupat... [desa, pagerwojo, letak, camat, budur, kabupat...
603 [bps, jawa, timur, tahun, kabupaten, tuban, ru... [bps, jawa, timur, tahun, kabupaten, tuban, ru...
385 [pengolahan, air, limbah, sesuai, peruntukanny... [olah, air, limbah, sesuai, untuk, untuk, awal...
11 [rumah, sakit, universitas, tangjungpura, ponti... [rumah, sakit, universitas, tangjungpura, ponti...
179 [keberadaan, ruang, terbuka, hijau, rth, kota,... [ada, ruang, buka, hijau, rth, kota, kendari, ...

```

Gambar 7. Hasil Remove Symbol

Hasil preprocessing dihasilkan menjadi seperti gambar berikut.

```

Gedung Fakultas Terpadu merupakan gedung berti... gedung fakultas padu gedung tingkat lantai kam...
Peningkatan konsumsi kopi di masyarakat kedai kopi da... tingkat konsumsi kopi di masyarakat kedai kopi da...
Bangkit Listrik Tenaga Uap (PLTU) milik PT ... bangkit listrik tenaga uap pltu milik pt pjb m...
Urin manusia menjadi penyumbang nutrisi terbes... urin manusia sumbang nutrisi besar air limbah ...
Semakin banyak sampah yang tidak tereduksi di ... sampah reduksi tps r akibat tambah anggaran biay...
Penggunaan vegetasi merupakan cara mengurangi ... guna vegetasi kurang polutan karbon dioksida c...
Pada perkembangan industri yang semakin tinggi... kembang industri kualitas lingkung cenderung a...
Desa Ngepung yang terletak di Kecamatan Kedame... desa ngepung letak camat dame kabupaten gresik...
Timbulan sampah Kota Sukabumi sebesar 171 ton/... timbul sampah kota sukabumi tonhari sampah ola...
Kota Surabaya adalah salah satu kota dengan la... kota surabaya salah kota laju tumbuh duduk pad...
Jalan Tol Krian - Legundi - Bunder merupakan p... jalan tol krian legundi bunder proyeksional ...
Surabaya's strategic location in the heart of ... surabayas strategic location in the heart of i...
Sungai Buntung Sidoarjo berada di kawasan perm... sungai buntung sidoarjo kawasan mukim padat du...
IPAL domestik di Kota Malang berjumlah 102 ban... ipal domestik kota malang bangun sebar camat i...
PT. SUPRAMA merupakan salah satu industri yang... pt suprama salah industri kecimpung bidang bua...
Kegiatan pembangunan yang diiringi dengan perk... giat bangun iring kembang teknologi sektor tra...
Pengelolaan sampah adalah bagian penting dari ... kelola sampah daerah mukim lingkung sehat sela...
Sampah organik merupakan jenis sampah yang mud... sampah organik jenis sampah mudah busuk potens...
Pesantren Modern Al-Amanah Junwangi merupakan ... pesantren modern alamanah junwangi salah pond...
PT X Indonesia merupakan sebuah perusahaan yan... pt x indonesia usaha lokasi kawasan industri t...

```

Gambar 8. Hasil Preprocessing

3.3. Proses LDA

Setelah preprocessing dilakukan menggunakan metode alokasi latent dirichlet. Hal ini digunakan untuk mengetahui kata topik mana yang mewakili masing-masing topik, dan kemudian melakukan pengelompokan dengan masing-masing topik. Pengelompokkan menjadi topik 0, topik 1, topik 2, topik 3, topik 4, topik 5, topik 6, dan topik 7.

Topik 0:

- co (0.0842)
- emisi (0.0268)
- hasil (0.0142)
- rth (0.0129)
- green (0.0127)
- tahun (0.0123)
- emissions (0.0121)
- vegetasi (0.0115)
- rata (0.0114)
- vegetation (0.0107)
- konsentrasi (0.0106)
- daya (0.0104)
- sebesar (0.0100)
- temperatur (0.0097)
- perhitungan (0.0092)

Topik 1:

- and (0.0170)
- of (0.0131)
- variasi (0.0125)
- konsentrasi (0.0111)
- hari (0.0106)
- reaktor (0.0098)

Gambar 9. Hasil Topic Modelling

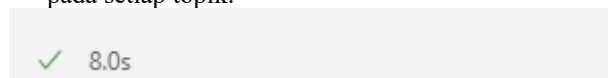
Setelah dilakukannya topik modelling penulis melakukan untuk mencari dominan topik pada setiap teks, Berikut hasilnya

Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords
0	0	3	0.722900 logam, air, tumbuhan, minum, daun, metode, tan...
1	1	4	0.916646 sampah, sebesar, limbah, pengelolaan, data, ha...
2	2	5	0.933189 the, of, and, to, in, is, that, water, with, on
3	3	1	0.779559 and, of, variasi, konsentrasi, hari, reaktor, ...
4	4	4	0.879843 sampah, sebesar, limbah, pengelolaan, data, ha...
5	5	0	0.957228 co, emisi, hasil, rth, green, tahun, emissions...
6	6	5	0.920106 the, of, and, to, in, is, that, water, with, on

Gambar 10. Hasil Dominan Topic

3.4. Nilai coherence

Dalam penelitian ini, model topik menggunakan metode Latent Dirichlet Allocation (LDA) dievaluasi berdasarkan nilai koherensi. Nilai koherensi digunakan untuk menilai seberapa baik topik-topik yang didapatkan oleh model dapat merepresentasikan korelasi semantik antar kata pada setiap topik.



Coherence Score: 0.4885446223476974

Gambar 11. Hasil Nilai Coherence

Pada hasil analisis ini, nilai koherensi sebesar ≈ 0.4885 menunjukkan bahwa model topik yang dibangun mempunyai tingkat koherensi yang cukup baik, meskipun belum mencapai ambang umum 0.5 yang biasanya dianggap sebagai indikator topik-topik yang bermakna. Hal ini masih dapat diterima mengingat kompleksitas data dan keterbatasan jumlah korpus. Secara umum, nilai koherensi yang mendekati atau lebih besar dari 0.5 dianggap mencerminkan kualitas model topik yang baik dalam menghasilkan topik-topik yang bermakna[14]. Meskipun demikian, interpretasi hasil tetap memerlukan validasi terhadap kata-kata pada masing-masing topik untuk memastikan relevansi dengan domain studi, yaitu skripsi Prodi Teknik Lingkungan dari empat kampus: UINSA, ITS, dan UNUSIDA.

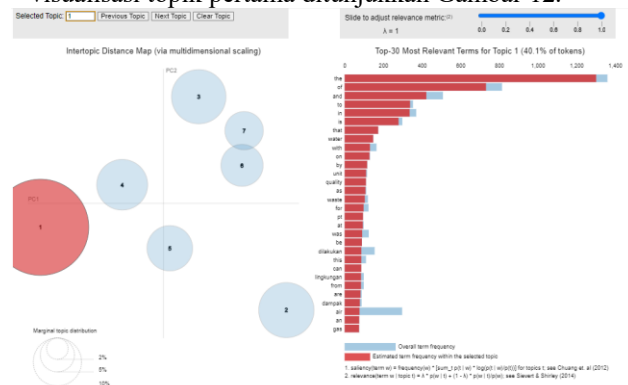
Koherensi model yang relatif baik ini mencerminkan keberhasilan preprocessing data, seperti tokenisasi, penghapusan kata-kata tidak penting (stopwords), serta penggunaan leksikon khusus untuk teknik lingkungan. Proses ini

mendukung model LDA dalam mengidentifikasi pola-pola tema utama dari korpus data skripsi.

3.5. Visualisasi

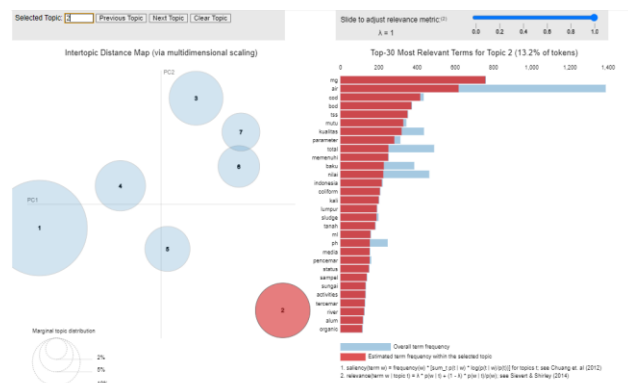
Selanjutnya adalah tahap visualisasi agar dapat dengan mudah dianalisis. Visualisasi menggunakan pyLDAvis dirancang untuk menafsirkan topik dalam model topik yang sesuai dengan kumpulan data teks. Visualisasi PyLDAvis akan selalu menampilkan 30 term pertama dalam setiap topik[15].

Visualisasi dilakukan dengan menganalisis kata di setiap kelompok topik diterapkan dengan PyLDAvis. Analisis topik dilakukan dengan memeriksa setiap kelompok dalam model LDA yang dihasilkan. Untuk menemukan jumlah kelompok, metode nilai koherensi digunakan untuk menetapkan jumlah kelompok yang paling sesuai. Setelah jumlah kelompok diketahui, model LDA kemudian akan dibuat dengan parameter jumlah topik berdasarkan hasil nilai koherensi. Hasil visualisasi topik pertama ditunjukkan Gambar 12.



Gambar 12. Hasil Visualisasi Topik 1

Berdasarkan gambar 12. Output topik 1 membahas gas, air, dan dampak lingkungan. Sumbu x dan y diwakili oleh pc1 dan pc2 di sisi kiri gambar. Pada grafik sisi kanan, lingkaran nomor 1 berwarna merah menunjukkan topik 1, yang berarti 40,1% dari semua dokumen yang menunjukkan topik 1 relevan. Pada grafik, topik 1 menunjukkan jumlah kata yang muncul, dan seluruh dokumen ditandai dengan warna biru.



kata yang muncul, dan seluruh dokumen ditandai dengan warna biru.

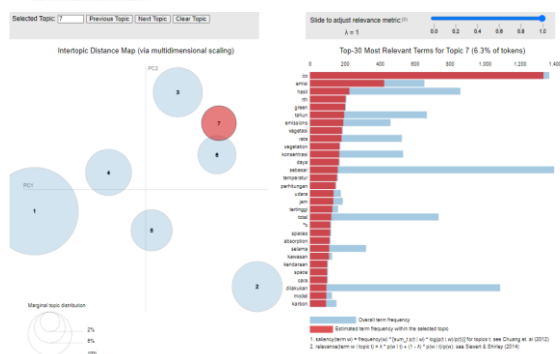
[illegible]

Berdasarkan gambar 16. Output topik satu mencakup perencanaan, limbah, pipa, dan penduduk. Sumbu x dan y diwakili oleh pc1 dan pc2 di sisi kiri gambar. Pada grafik sisi kanan, lingkaran angka 1 berwarna merah menunjukkan topik 1. Ini berarti bahwa 8,9% dari semua dokumen yang menunjukkan topik 1 relevan. Pada grafik, topik 1 menunjukkan jumlah kata yang muncul, dan seluruh dokumen ditandai dengan warna biru.

[illegible]

Berdasarkan gambar 17. output topik 1 membahas mengenai logam, warna, tumbuhan, dan minyak. Sumbu x dan y diwakili oleh pc1 dan pc2 di sisi kiri gambar. Pada grafik sisi kanan, lingkaran angka 1 berwarna merah menunjukkan topik 1. Ini berarti bahwa 7,6% dari semua dokumen yang menunjukkan topik 1 adalah relevan. Pada grafik, topik 1 menunjukkan jumlah kata yang muncul, dan seluruh dokumen ditandai dengan warna biru.

Jurnal Teknologi Dan Sistem Informasi Bisnis – JTEKSIS Vol.7 No.4 Oktober 2025



Gambar 18. Hasil Visualisasi Topik 7

Berdasarkan gambar 18. Output topik 1 mencakup suhu, udara, area, dan kendaraan. Sumbu x dan y diwakili oleh pc1 dan pc2 di sisi kiri gambar. Pada grafik sisi kanan, lingkaran angka 1 berwarna merah menunjukkan topik 1, yang berarti 6,3% dari semua dokumen yang mewakili topik 1 relevan. Pada grafik, topik 1 menunjukkan jumlah kata yang muncul, dan seluruh dokumen ditandai dengan warna biru.

Analisis menghasilkan beberapa topik utama dari skripsi Teknik Lingkungan. Temuan ini sejalan dengan penelitian-sejenis yang menggunakan LDA pada abstrak tugas akhir, misalnya penelitian pada mahasiswa Fakultas Informatika ITTP yang menghasilkan 8 topik dengan nilai coherence $\approx 0,4468$ [16]. Teknik pemilihan jumlah topik melalui evaluasi nilai koherensi juga digunakan dalam literatur terkini yang membahas mekanisme optimasi jumlah topik LDA [17].

SIMPULAN

Pemodelan topik menggunakan metode *Latent Dirichlet Allocation* (LDA) pada dataset berisi 644 abstraksi skripsi mahasiswa Program Studi Teknik Lingkungan selama empat tahun terakhir berhasil mengidentifikasi tujuh topik utama dengan nilai coherence sebesar 0,4885. Nilai tersebut menunjukkan bahwa model memiliki konsistensi yang cukup baik dalam menggambarkan keterkaitan antar topik [18]. Berdasarkan analisis terhadap tiga *term* teratas dari setiap kelompok topik, diperoleh bahwa penelitian mahasiswa Teknik Lingkungan di Jawa Timur berfokus pada isu-isu seperti pengelolaan limbah, kualitas air, pengendalian pencemaran udara, serta penerapan teknologi pengolahan lingkungan. Namun demikian, masih ditemukan sejumlah *term* yang kurang relevan dengan konteks tugas akhir, seperti “lingkungan”, “rumah”, “dana”, “belajar”, dan “air”. Hal ini menunjukkan bahwa proses pembersihan data perlu ditingkatkan agar hasil pemodelan lebih representatif terhadap substansi penelitian sebenarnya [19].

Penelitian ini memberikan kontribusi nyata dalam memetakan tren penelitian mahasiswa Teknik Lingkungan, yang dapat dimanfaatkan sebagai dasar pengambilan keputusan akademik, penyusunan kurikulum, maupun perencanaan arah penelitian program studi [20]. Untuk penelitian selanjutnya, peningkatan kualitas *data preprocessing* dan penyesuaian parameter LDA (melalui *hyperparameter tuning*) diharapkan dapat meningkatkan akurasi model [21]. Selain itu, perbandingan dengan metode pemodelan lain seperti *Non-negative Matrix Factorization* (NMF) dan *BERTopic* perlu dilakukan untuk memperoleh gambaran yang lebih komprehensif mengenai struktur tematik penelitian [22]. Hasil analisis ini juga berpotensi digunakan sebagai landasan dalam penyusunan *roadmap* riset kampus, sehingga arah penelitian mahasiswa dapat semakin selaras dengan isu lingkungan terkini dan kebutuhan pengembangan ilmu di bidang Teknik Lingkungan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih atas kesempatan yang telah diberikan untuk menyelesaikan penelitian ini. Tentu juga tak lupa kepada Universitas Islam Negeri Sunan Ampel yang telah memberikan pengarahan, kesempatan, serta dukungannya kepada penulis sehingga penulis dapat menyelesaikan penelitian ini. Selain itu, penulis berharap hasil penelitian ini akan bermanfaat untuk masyarakat dan memberikan informasi baru.

DAFTAR PUSTAKA

- [1] Universitas Pembangunan Nasional Veteran Jawa Timur, *Analisis dampak pencemaran lingkungan pada IPAL PT. SIER Surabaya dengan metode life cycle assessment (LCA)*, Skripsi, Prodi Teknik Lingkungan, 2021.
- [2] Universitas Islam Negeri Sunan Ampel Surabaya, *Analisis jejak karbon (carbon footprint) pada aktivitas permukiman dan kualitas air di Jawa Timur*, Skripsi, Prodi Teknik Lingkungan, 2022.
- [3] Universitas Pembangunan Nasional Veteran Jawa Timur, *Potensi produksi gas metana dari landfilling di TPA Bengkala Kabupaten Buleleng dengan kombinasi model LandGEM, IPCC, dan LCA*, Skripsi, Prodi Teknik Lingkungan, 2023.
- [4] G. Supiadin and A. D. Laksito, “Evaluating LDA and LSA for topic modeling in the Indonesian natural disaster,” *Int. J. Comput. Sci. (IJCS)*, 2023.
- [5] H. Hairani, M. Janhasmadja, A. Tholib, J. X. Guterres, and Y. Ariyanto, “Thesis topic modeling study: Latent Dirichlet allocation (LDA) and machine learning approach,” *Int. J. Eng. Comput. Sci. Appl. (IJECSA)*, Universitas Bumigora, 2024.
- [6] N. Idrus, *Analisis topic modeling menggunakan LDA dan HDP dengan word embedding dan clustering*, Tesis, Universitas Gadjah Mada, 2024.
- [7] A. Siino, S. Brasini, dan D. Aiello, “Natural Language Processing in the Age of Large Language Models: A Survey on Pre-processing Techniques and Challenges,” *Information Processing & Management*, vol. 61, no. 6, p. 103678, 2024.
- [8] R. A. Berutu, R. Mayasari, dan D. S. Purnamasari, “Text Pre-processing for Indonesian Language in Sentiment Analysis,” *Jurnal Infotel*, vol. 15, no. 3, pp. 161–170, 2023.
- [9] Y. Farea, H. A. Aboalsamh, dan M. A. Bamahdi, “An Enhanced Topic Modeling Approach Using LDA with Optimal

Hyperparameters,” *Engineering Applications of Artificial Intelligence*, vol. 132, p. 107289, 2024.

[10] Casualty Actuarial Society, “Latent Dirichlet Allocation Topic Modeling in Python,” *CAS Research Paper*, 2023. [Online]. Available:

https://www.casact.org/sites/default/files/2023-04/iCAS-4_Latent_Dirichlet_Allocation_Topic_Modeling_in_Python.pdf

[11] D. M. Blei, A. Y. Ng, dan M. I. Jordan, “Latent Dirichlet Allocation Revisited with Modern Applications,” *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–40, 2022.

[12] H. U. Onah, C. J. C. Burges, dan R. S. W. Onah, “Interactive Topic Visualization for Large-Scale Text Data,” *arXiv preprint*, arXiv:2207.14687, 2022.

[13] N. Novarian et al., “Topic Modeling Tugas Akhir Mahasiswa Fakultas Informatika,” *Ledger*, vol. 2, no. 1, Feb. 2023.

[14] A. Farea et al., “Investigating the optimal number of topics by advanced text segmentation for coherence optimization,” *Elsevier*, 2024.

[15] H. U. Onah et al., “Interactive Topic Visualization for Large-Scale Text Data,” *arXiv preprint*, 2022.

[16] N. Novarian et al., “Topic Modeling Tugas Akhir Mahasiswa Fakultas Informatika,” *Ledger*, vol. 2, no. 1, Feb. 2023.

[17] A. Farea et al., “Investigating the optimal number of topics by advanced text segmentation for coherence optimization,” *Elsevier*, 2024.

[18] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent Dirichlet Allocation*. *Journal of Machine Learning Research*, 3, 993–1022.

[19] Syahputra, M., & Rahman, F. (2022). *Improving Topic Modeling Performance through Enhanced Text Preprocessing in Indonesian Research Abstracts*. *Indonesian Journal of Information Systems*, 5(2), 101–112.

[20] Lee, J., & Chen, S. (2021). *Mapping Research Trends Using LDA: A Case Study in Environmental Engineering*. *Environmental Informatics Journal*, 18(4), 245–258.

[21] Raharja, D., & Sari, N. (2023). *Optimization of LDA Hyperparameters for Scientific Text Analysis*. *Journal of Data Science and Applications*, 9(1), 55–67.

[22] Kaur, P., & Singh, M. (2024). *Comparative Study of Topic Modeling Techniques: LDA, NMF, and BERTopic*. *Applied Computational Intelligence and Soft Computing*, 2024, 1–11.