

# Identifikasi Potensi Topik Penelitian Baru: Analisis Judul Skripsi Teknik Informatika (2022-2024) menggunakan Latent Dirichlet Allocation (LDA)

Angdy Erna<sup>\*1</sup>, Abdul Kadir Jailani<sup>2</sup>

<sup>1</sup>Sistem Informasi, Universitas Dipa, Makassar

<sup>2</sup>Teknik Informatika, Universitas Dipa, Makassar

e-mail: <sup>\*1</sup>angdy@undipa.ac.id, <sup>2</sup>akjailani@undipa.ac.id

## Abstrak

Penelitian ini bertujuan mengidentifikasi potensi topik penelitian baru di bidang Teknik Informatika dengan menggunakan teknik pemrosesan bahasa alami. Analisis dilakukan terhadap judul skripsi yang dihasilkan dari tahun 2022 hingga 2024. Metode Latent Dirichlet Allocation (LDA) diterapkan untuk mengidentifikasi tema-tema latendalam judul skripsi. Hasil penelitian menunjukkan bahwa metode ini menghasilkan 13 topik yang relevan dengan perkembangan terkini di bidang Teknik Informatika diantaranya data mining, machine learning, sistem berbasis web, android, analisis sentimen, IoT dll. Nilai koherensi yang diperoleh adalah 0.3. Selain itu, penelitian ini juga berhasil mengidentifikasi beberapa topik penelitian baru yang belum banyak diteliti sebelumnya, seperti interseksi antara potensi wisata dengan AI untuk personalisasi rekomendasi, Pengembangan sistem rekomendasi mata kuliah berbasis AI yang mempertimbangkan minat, kemampuan, dan tujuan karier mahasiswa, dll. Temuan ini diharapkan dapat memberikan wawasan berharga bagi akademisi dan praktisi dalam mengarahkan arah penelitian di masa mendatang, serta membantu institusi pendidikan dalam merancang kurikulum yang relevan dan adaptif terhadap perkembangan teknologi dan kebutuhan industri.

**Kata kunci :** Pemrosesan Bahasa Alami, Tema Laten, Topik Penelitian Baru, LDA, Teknik Informatika

## Abstract

*This research aims to identify potential new research topics in Informatics by employing natural language processing techniques. The analysis was conducted on thesis titles produced from 2022 to 2024. The Latent Dirichlet Allocation (LDA) method was applied to identify latent topics within the thesis titles. The research results indicate that this method yields 13 topics relevant to current developments in Informatics, such as data mining, machine learning, web-based systems, Android, sentiment analysis, IoT, etc. The coherence score is 0.3. Furthermore, this research successfully identifies several new research topics that have not been widely explored previously, such as the intersection of tourism potential with AI for personalized recommendations and the development of AI-based course recommendation systems that consider students' interests, abilities, and career goals. These findings are expected to provide valuable insights for academics and practitioners in directing future research and assist educational institutions in designing relevant and adaptive curricula to meet technological advancements and industry needs.*

**Keywords:** Natural Language Processing, Latent Themes, Novel Research Topics, LDA, Informatics

## 1. Pendahuluan

Teknik Informatika (TI) adalah bidang yang berkembang pesat, didorong oleh kemajuan teknologi yang terus menerus dan permintaan yang semakin meningkat untuk solusi inovatif terhadap masalah-masalah kompleks. Oleh karena itu, mengidentifikasi topik penelitian yang sedang berkembang sangat penting bagi institusi akademik dan praktisi industri. Judul skripsi merupakan representasi ringkas dari suatu penelitian, mencerminkan pengetahuan dan keterampilan mahasiswa serta memberikan wawasan tentang arah penelitian di suatu institusi. Namun, dalam praktiknya, skripsi yang diajukan oleh mahasiswa terkadang tidak sesuai dengan deskripsi program studi ataupun konsentrasi [1], [2]. Analisis terhadap judul skripsi dapat memberikan wawasan berharga mengenai tren dan perubahan minat penelitian, mengungkapkan potensi area studi baru yang sesuai dengan perkembangan teknologi terkini dan kebutuhan masa depan. Dengan menganalisis judul skripsi, kita dapat pula memperoleh wawasan tentang arah penelitian yang diambil oleh lulusan dan mengevaluasi efektivitas program studi dalam mempersiapkan mahasiswa untuk karier masa depan mereka.

Penelitian ini bertujuan untuk mengidentifikasi topik penelitian yang dominan di bidang teknik informatika Universitas Dipa periode 2022-2024. Program Studi Teknik Informatika Universitas Dipa Makassar bertujuan menjadi yang unggul dan terdepan dalam bidang Teknologi Informasi dan Komputer pada tahun 2030. Program ini berkomitmen untuk menyediakan tenaga yang kompeten yang dapat menjawab tantangan perkembangan Teknologi Informasi dan Komputer dengan integritas dan berwawasan technopreneurship. Melalui pendidikan tinggi, program studi ini menghasilkan lulusan berkualitas di bidang Teknologi Informasi dan Komunikasi, berwawasan technopreneurship, serta siap bersaing secara lokal dan global. Selain itu, program studi ini juga berfokus pada menghasilkan lulusan yang mampu bersaing dalam menghadapi tantangan global di bidang penerapan Sistem Cerdas dan Jaringan Komputer Nirkabel.

Untuk memahami lebih dalam mengenai perkembangan topik penelitian Teknik Informatika di Universitas Dipa, diperlukan analisis yang sistematis terhadap sejumlah besar data. Salah satu pendekatan yang efektif adalah dengan menggunakan metode Latent Dirichlet Allocation (LDA). LDA adalah salah satu metode pemodelan topik dalam pemrosesan bahasa alami (Natural Language Processing =NLP) yang digunakan untuk mengidentifikasi tema-tema tersembunyi dalam kumpulan data teks yang besar. LDA bekerja dengan cara mengelompokkan kata-kata dalam dokumen ke dalam topik-topik yang tersembunyi, memungkinkan peneliti untuk melihat distribusi topik dalam kumpulan dokumen tersebut. Blei et al. [3] memperkenalkan LDA sebagai metode yang efektif untuk menemukan struktur tematik tersembunyi dalam korpus teks yang besar. Dengan menggunakan LDA, peneliti dapat mengidentifikasi distribusi topik dalam kumpulan judul skripsi, memberikan gambaran mengenai tema-tema dominan dan evolusi minat penelitian dari waktu ke waktu.

Beberapa studi telah menggunakan teknik pemodelan topik untuk menganalisis tren penelitian dan mengidentifikasi topik-topik yang sedang berkembang. Solihati I, et al. [4] menggunakan Graph Convolutional Network (GCN) untuk mengklasifikasi dan menganalisis tren skripsi mahasiswa TI. Hasil klasifikasi akurasi sebesar 68.25% dan tren topik yang diperoleh adalah matematika, statistika dan komputer arsitektur. Model ini sangat tergantung pada struktur graf, sehingga tidak dapat secara langsung digunakan pada data tanpa struktur graf. Utami N, et al. [5] mengelompokkan topik dokumen penelitian para dosen sebagai acuan dalam merancang roadmap penelitian perguruan tinggi menggunakan K-means dan Cosine Similarity. Terdapat 6 kluster yang dihasilkan yaitu Pengembangan dan Evaluasi Sistem Informasi, E- Government, Data Mining, Teknologi Pendidikan, Machine Learning/Artificial Intelligence, serta Manajemen dan Bisnis. Namun, K-means melakukan hard clustering dimana setiap judul penelitian hanya terkait dengan satu topik. Model kluster lainnya yaitu Hierarchical Dirichlet Process (HDP) mampu mengatasi hal tersebut. Erreza M, et al. [6] menggunakan HDP untuk menganalisis tren topik penelitian Jurnal Teknologi Informasi dan Robotika (JIFTI) tahun 2019-2022. Model ini dapat mengadaptasi jumlah topik sesuai dengan kompleksitas dataset juga dapat menangani situasi di mana dokumen memiliki campuran topik yang kompleks dan tidak terbatas [7], tetapi interpretasinya bisa lebih rumit karena adanya hierarki dan jumlah topik yang dinamis.

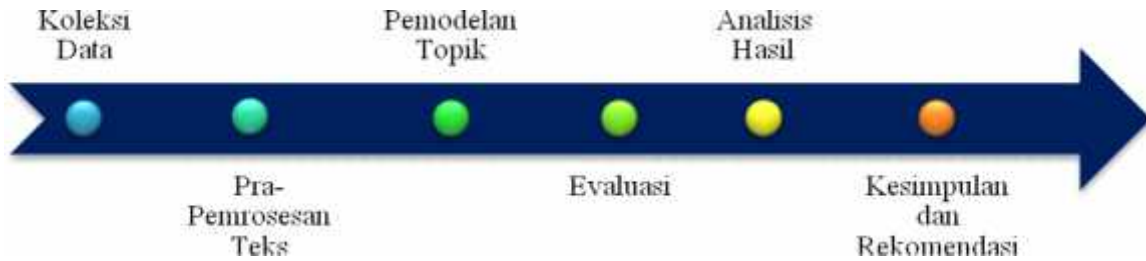
LDA adalah metode pemodelan topik yang sangat populer digunakan dalam analisis tema pada dokumen teks [8]. Keberhasilannya LDA sangat bergantung pada pemilihan jumlah topik yang tepat. Salah satu cara mengatasinya adalah dengan Iterative Refinement menggunakan metrik evaluasi. Nastiti et al. [9] telah berhasil menggunakan LDA untuk mengidentifikasi tren topik penelitian di bidang ilmu komputer secara umum. Namun, penelitian tersebut belum spesifik mengkaji judul skripsi di konteks institusi tertentu. Novarian et al. [10], Kamdan et al. [2] dan Titiana et al. [11] mengkaji topik-topik skripsi di lingkungan institusi, namun belum membandingkan hasil penelitian dengan penelitian sebelumnya. Penelitian ini berfokus untuk menemukan tren penelitian pada judul skripsi di Universitas Dipa. Selain itu, penelitian ini juga akan membandingkan hasil analisis dengan penelitian sebelumnya untuk mengungkap celah topik penelitian. Perbandingan ini diharapkan dapat memberikan gambaran yang lebih komprehensif tentang tren penelitian di bidang teknik informatika, memberikan kontribusi bagi pengembangan kurikulum, identifikasi peluang penelitian baru, serta mendukung pencapaian visi Universitas Dipa untuk menjadi pusat unggulan dalam bidang teknologi informasi.

## **2. Metode Penelitian**

### **2.1 Tahapan Penelitian**

Gambar 1 menunjukkan tahapan dalam penerapan LDA untuk analisis topik, dimulai dengan koleksi data, di mana data judul skripsi dikumpulkan untuk dianalisis. Langkah berikutnya adalah pra-pemrosesan teks, yang melibatkan pembersihan dan normalisasi teks agar siap untuk tahap selanjutnya. Kemudian, dilakukan pemodelan topik menggunakan LDA untuk mengidentifikasi topik laten dalam data. Evaluasi dilakukan untuk menilai kualitas topik yang dihasilkan, memastikan bahwa topik

yang ditemukan relevan dan bermakna. Hasil dari pemodelan ini kemudian dianalisis dalam tahap analisis hasil, di mana topik-topik yang diidentifikasi dievaluasi berdasarkan relevansi dan kemunculannya. Terakhir, kesimpulan dan rekomendasi disusun berdasarkan temuan, memberikan panduan bagi penelitian lebih lanjut atau pengembangan kurikulum yang sesuai.



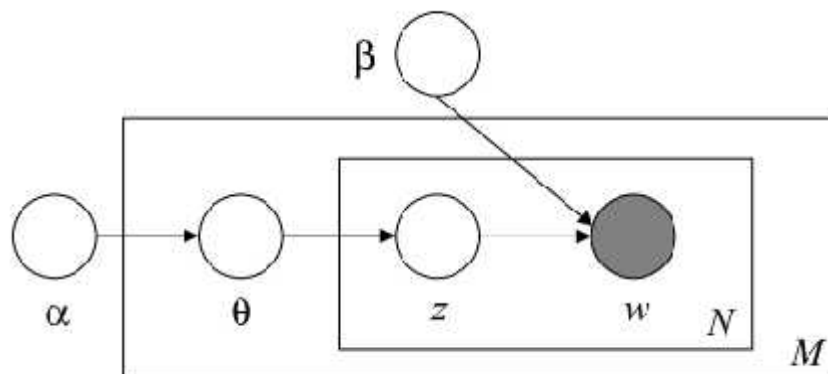
Gambar 1. Tahapan Penelitian

## 2.2 Koleksi Data dan Pra-Pemrosesan Teks

Data diperoleh dari judul-judul penelitian yang ditulis oleh mahasiswa program studi Teknik Informatika Universitas Dipa Makassar tahun 2022 sampai tiga bulan pertama tahun 2024 dan berjumlah 538 judul penelitian. Fokus pada data judul skripsi berdasarkan acuan kurikulum prodi akan memberikan gambaran yang lebih relevan dan terkini tentang tren penelitian di bidang Teknik Informatika. Dengan periode yang lebih singkat, proses analisis akan lebih konsisten, efisien, spesifik, dan fokus pada identifikasi topik penelitian baru yang sesuai dengan tantangan dan peluang masa kini. Untuk pra-pemrosesan teks, peneliti memanfaatkan *Natural Language Tool Kit* (NLTK) yaitu *tokenizing*, *filtering* dan untuk *stemming*.

## 2.3 Latent Dirichlet Allocation

*Latent Dirichlet Allocation* (LDA) adalah model generatif probabilistik yang digunakan untuk melakukan pemodelan topik dalam teks. Algoritma ini pertama kali diperkenalkan oleh Blei et al. [3]. LDA berdasarkan pada asumsi dasar bahwa setiap dokumen dalam koleksi dokumen terdiri dari kombinasi tertentu dari topik-topik yang ada, dan setiap topik diwakili oleh distribusi kata-kata tertentu.



Gambar 2. Representasi Grafis Model LDA [3]

LDA dirumuskan sebagai berikut:

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta \quad (1)$$

Dimana  $M$  adalah jumlah dokumen;  $N$  adalah jumlah kata dalam dokumen tertentu;  $i$  adalah *dirichlet prior* pada sebaran topik per dokumen;  $j$  adalah *dirichlet prior* pada sebaran kata per topik;  $i$  adalah sebaran topik untuk dokumen  $i$ ;  $z_{ij}$  adalah topik untuk kata ke- $j$  pada dokumen  $i$ ;  $w_{ij}$  adalah kata spesifik.

**2.4 Metrik Evaluasi Koherensi**

Metrik evaluasi koherensi dalam pemodelan topik, khususnya LDA, sangat penting untuk mengukur seberapa baik topik-topik yang ditemukan oleh model mewakili tema-tema yang sebenarnya ada dalam data. Metrik ini membantu untuk menilai kualitas model, memilih jumlah topik yang optimal, dan memvalidasi hasil analisis. Penelitian ini memanfaatkan metrik koherensi  $c_v$  yang tersedia di pustaka *Gensim Python*. Koherensi  $c_v$  mempertimbangkan kesamaan antar kata dan distribusi probabilitas kata untuk mengukur konsistensi antar topik. Metrik ini didasarkan pada kesamaan antar pasangan kata, sering kali menggunakan model ruang vektor untuk menganalisis hubungan antara setiap kata.  $C_v$  biasanya dinyatakan sebagai nilai antara 0 dan 1; semakin mendekati 1, semakin tinggi konsistensi topiknya[9].

**3. Hasil dan Pembahasan**

**3.1 Koleksi Data**

Data orisinal skripsi TI terdiri dari 9 atribut tersaji pada gambar 3. Peneliti mengeliminasi atribut-atribut yang tidak relevan dengan tujuan penelitian dan menyisakan 2 atribut yaitu *JudulSkripsi* dan *TahunTerbit*. Gambar 4 menunjukkan data *csv* judul penelitian tahun 2022-2024 yang berjumlah 538.

Gambar 3. Data Judul Skripsi TI tahun 2022-2024

	Judul	Tahun
0	Perancangan Sistem Data Center Berbasis Mobile...	2022
1	Implementasi Algoritma Sequential Search pada ...	2022
2	Implementasi Aplikasi Pengolaan Data Travel Um...	2022
3	Analisis Faktor Pemilihan Aplikasi Media Sosia...	2022
4	Perancangan dan Implementasi Algoritma Elgamal...	2022
...	...	...
533	Implementasi Metode Local Binary Pattern Histo...	2024
534	Analisis Penyakit Tanaman Sawi Menggunakan Met...	2024
535	Analisis Perbandingan Sistem Prediksi Masa Stu...	2023
536	Pengembangan Aplikasi Edukasi untuk Pengenalan...	2024
537	Klasifikasi Citra Daun Tanaman Herbal Mengguna...	2024

[538 rows x 2 columns]

Gambar 4. Data *csv* Judul Skripsi TI tahun 2022-2024

### 3.2 Preprocessing Text

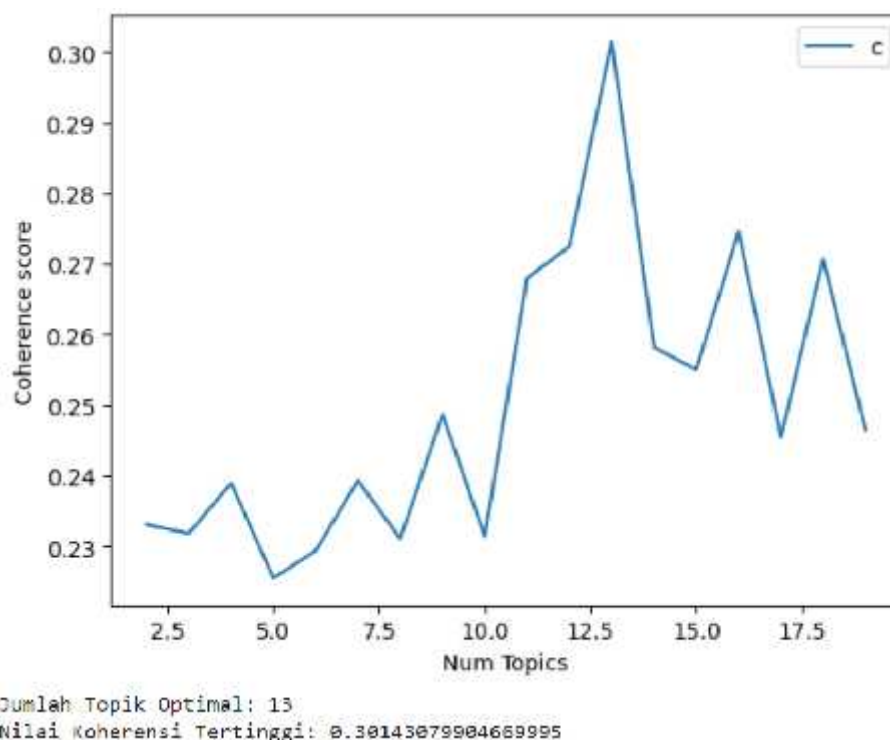
(1) *Tokenizing*. Pada tahap ini, teks yang telah melewati tahap Case Folding akan dilakukan proses pemecahan perkata menggunakan fungsi `.word_tokenize()` pada library NLTK. Selain itu pada tahap ini juga akan dilakukan proses removing number, whitespace dan punctuation (tanda baca).

(2) *Filtering*. Tahap ini bertujuan untuk mengambil kata-kata penting pada *tokens* yang dihasilkan oleh proses sebelumnya. Kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna disebut *Stopword*. Contoh *stopword* dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Fungsi `.stopword()` pada library NLTK untuk mendapatkan list *Indonesian stopwords*.

(3) *Stemming*. Tahap ini akan menghilangkan *suffix* dan *prefix* pada token/kata, sehingga sebuah kata yang memiliki *suffix* maupun *prefix* akan kembali ke bentuk dasarnya. Fungsi yang digunakan adalah `PorterStemmer()` dari NLTK

### 3.3 Proses LDA

Langkah selanjutnya adalah pemodelan LDA. Untuk membangun model LDA, penelitian ini memanfaatkan pustaka *Gensim* dengan fungsi `LdaModel()`, sedangkan untuk mendapatkan nilai koherensi, digunakan fungsi `CoherenceModel()`. Nilai koherensi diperoleh setelah model dibuat. Fungsi `CoherenceModel()` diterapkan untuk mengevaluasi hasil model topik. Jumlah topik dengan nilai koherensi tertinggi akan dipilih sebagai jumlah topik yang paling tepat. Dalam eksperimen ini, peneliti menetapkan rentang jumlah topik dari 2 hingga 20 topik untuk mendapatkan jumlah topik terbaik. Hasil perhitungan nilai koherensi dapat dilihat pada Gambar 5, terlihat bahwa nilai koherensi tertinggi dicapai ketika jumlah topik sama dengan 13. Oleh karena itu, dapat disimpulkan bahwa jumlah topik yang paling sesuai adalah 13 topik.

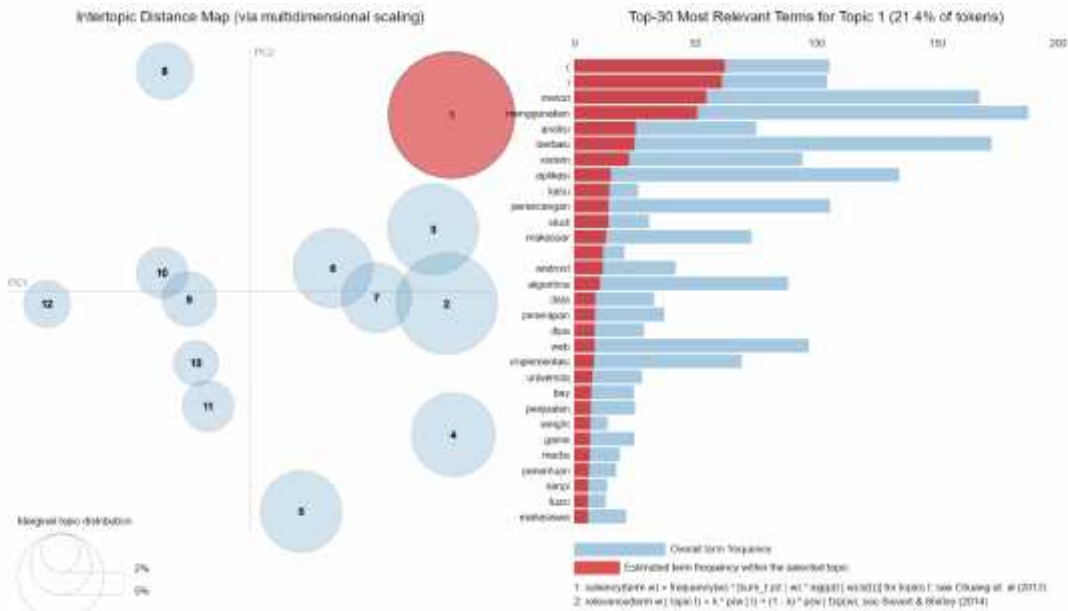


Gambar 5. Nilai Koherensi DataPeriode 2022-2024

### 3.3 Visualisasi

Untuk memudahkan proses interpretasi hasil, dilakukan visualisasi topik menggunakan *pyLDAvis* dan *wordcloud*. *PyLDAvis* adalah visualisasi interaktif yang ditulis oleh Sievert et al. [12], menampilkan 30 kata penting dalam korpus. Di panel kiri, ukuran lingkaran menunjukkan seberapa sering topik muncul dalam korpus. Selain itu, panel ini juga menunjukkan seberapa dekat atau jauh topik-topik tersebut satu sama lain. Ini menunjukkan kemiripan atau perbedaan topik-topik tersebut. Lingkaran yang dekat satu sama lain menunjukkan topik-topik yang saling terkait. Panel kanan menampilkan informasi tentang frekuensi keseluruhan suatu istilah serta frekuensi istilah tersebut dalam topik yang sedang dipilih dalam bentuk *bar chart*. Istilah dengan bar yang panjang umumnya penting dalam topik tersebut. Sebagai

contoh, Gambar 6 menunjukkan bahwa kata-kata kunci seperti metode, sistem, analisis, aplikasi, perancangan, implementasi, android dan web merupakan istilah atau tema dari "aplikasi berbasis android atau web". Tema ini mendominasi sangat menonjol dalam topik pertama, yang menunjukkan bahwa topik utama yang dibahas dalam topik nomor satu adalah analisis dan perancangan sistem informasi.



Gambar 6. Visualisasi Topik Pertama dengan pyLDAvis

Peneliti juga memvisualisasikan kata-kata yang paling sering muncul dalam suatu topik menggunakan *wordcloud* (awan kata). Dalam awan kata, ukuran kata mencerminkan seberapa penting kata-kata tersebut dalam kumpulan teks. Semakin besar ukuran sebuah kata/istilah/frasa, semakin sering kata/istilah/frasa tersebut muncul dalam teks. Awan kata di set berdasarkan top 5-20 frekuensi kemunculan kata. Gambar 7 menunjukkan contoh awan kata dari 2 topik. Berdasarkan ukuran istilah, awan kata topik 4 membahas relevansi topik pada pengembangan aplikasi berbasis teknologidalam berbagai konteks seperti pendidikan, bisnis, dan studi kasus di wilayah tertentu. Awan kata topik 8 berfokus pada penggunaan algoritma machine learning seperti Support Vector Machine (SVM) dan Naïve Bayes untuk analisis sentimen.



Gambar 7. Awan Kata Topik 4 dan 8 (Top 5-20 frequency)

### 3.4 Interpretasi Hasil dan Perbandingan

Interpretasi topik dalam konteks hasil LDA melibatkan analisis kata-kata kunci yang muncul dalam setiap topik yang diidentifikasi oleh model. Berdasarkan kata-kata kunci ini, interpretasi dapat memberikan wawasan tentang tema utama dari masing-masing topik. Tabel 1 menunjukkan hasil interpretasi 13 topik hasil LDA.

TOPIK	KATA KUNCI RELEVAN	INTERPRETASI
TOPIK 1	METOD, MENGGUNAKAN, ANALISI, BERBASI, SISTEM, APLIKASI, KASU, PERANCANGAN, STUDI, MAKASSAR, ANDROID, ALGORITMA, DATA, PENERAPAN, DIPA, WEB, IMPLEMENTASI, UNIVERSITA	<b>PENGEMBANGAN APLIKASI ANDROID ATAU WEB</b>
TOPIK 2	METOD, MENGGUNAKAN, BERBASI, APLIKASI, PERANCANGAN, WEB, SISTEM, MAKASSAR, ALGORITMA, IMPLEMENTASI, PENJUALAN, ANALISI, TINGKAT KLASIFIKASI DATA PENERAPAN INFORMASI SERVIC PREDIKSI PRODUKSI KEPUASAN REGRESI WEBSIT UNIVERSITA DIPA PERBANDINGAN PENGGUNA	<b>DATA MINING UNTUK MANAJEMEN PENJUALAN DAN KEPUASAN PENGGUNA</b>
TOPIK 3	BERBASI WEB APLIKASI ) ( PERANCANGAN SERVIC MENGGUNAKAN METOD SISTEM MAKASSAR ANALISI IMPLEMENTASI ALGORITMA KASU STUDI : PENERAPAN INFORMASI MAHASISWA RANCANG BANGUN MEDIA TERHADAP SIMPL PERBANDINGAN UNIVERSITA DIPA ADDIT WEIGHT	<b>SISTEM DAN LAYANAN DI UNIVERSITAS</b>
TOPIK 4	BERBASI APLIKASI PERANCANGAN IMPLEMENTASI METOD WEB MENGGUNAKAN ALGORITMA MAKASSAR ANDROID GAME SISTEM PENGENALAN EDUKASI SULAWESI SELATAN PENJUALAN SERVIC	<b>GAME EDUKASI</b>
TOPIK 5	MENGGUNAKAN BERBASI APLIKASI ALGORITMA WEB PERANCANGAN SISTEM MAKASSAR IMPLEMENTASI RANCANG BANGUN MONITOR PENERAPAN SENSOR ANDROID METOD UNIVERSITA DIPA PARKIR EDUKASI BAHASA REKOMENDASI BUKU PENGEMBANGAN PRODUK BAY GAME ANALISI IOT KOTA	<b>MONITOR PARKIR BERBASIS SENSOR</b>
TOPIK 6	MENGGUNAKAN METOD BERBASI APLIKASI ANALISI WEB ANDROID RANCANG BANGUN MAKASSAR ALGORITMA SISTEM BAY PERANCANGAN PEMBELAJARAN ANAK SUPPORT INFORMASI VECTOR ONLIN NAÏV INTERAKTIF IMPLEMENTASI BARANG DINA PENYAKIT MODEL REALITI	<b>PEMBELAJARAN INTERAKTIF UNTUK ANAK</b>
TOPIK 7	METOD MENGGUNAKAN BERBASI PERANCANGAN SISTEM APLIKASI ) ( DATA IMPLEMENTASI ANDROID MAKASSAR THING OF INTERNET WEB LOKASI ALAT ALGORITMA REKOMENDASI BAY : PENYAKIT OUTDOOR UTARA JARAK ARDUINO E-COMMERC KHA COLLABOR.	<b>INTERNET OF THING (IOT)</b>
TOPIK 8	MENGGUNAKAN ANALISI METOD ALGORITMA WEB BERBASI VECTOR MACHIN SENTIMEN TERHADAP SUPPORT SISTEM ( PERBANDINGAN MEDIA ) SUHU DISTRIBUSI PROMOSI MANAJEMEN MOTOR BERDASARKAN NAÏV TIKTOK PREDIKSI BAY SERVIC KEPUTUSAN MODEL PENENTUAN	<b>ANALISIS SENTIMEN PROMOSI VIA MEDIA SOSIAL</b>
TOPIK 9	ALGORITMA APLIKASI MENGGUNAKAN BERBASI PERANCANGAN SISTEM PENERAPAN IMPLEMENTASI WEB UNIVERSITA DIPA MAKASSAR METOD CLUSTER SEARCH KLASIFIKASI ANALISI PENDUKUNG KEPUTUSAN SELATAN SULAWESI FUTSAL INVESTASI MULTI PROCESS RANCANG BANGUN CALON PENERIMA DATA	<b>KLASIFIKASI DAN ANALISIS KLASTER</b>
TOPIK 10	IMPLEMENTASI BERBASI MENGGUNAKAN ALGORITMA	<b>STUDI KASUS</b>

	MAKASSAR ANALISI PERANCANGAN METOD ANDROID APLIKASI DIPA UNIVERSITA SISTEM STUDI BERKAH PEKERJAAN RANDOM MANDIRI OTOMATI WEB BERDASARKAN NAİV PENENTUAN BAY KLASIFIKASI PEMBUATAN ANAK ) KASU	<b>UNIVERSITAS DIPA</b>
TOPIK 11	APLIKASI PERANCANGAN ALGORITMA METOD MENGGUNAKAN BERBASI MOBIL MAKASSAR WEB DINI GAME SISTEM IMPLEMENTASI PRESIDEN SEBAGAI PERBANDINGAN MAGANG CALON PROVINSI PT INFORMASI PENERAPAN ANALISI RANDOM MINE DATA DETEKSI WEIGHT ADDIT SIMPL	<b>SISTEM DETEKSI DINI DAN APLIKASI GAME MOBILE</b>
TOPIK 12	ANALISI ALGORITMA MENGGUNAKAN IMPLEMENTASI BERBASI DATA APLIKASI AKURASI SENTIMEN MAKASSAR JARINGAN PT WEB PERANCANGAN TOUR METOD BERDASARKAN RUNNER REDUD PROTOCOL CIVITA LIKERT MULTIVARIAT MESSAG ADAKAMI POLA K-NN TANGGAP WHATSAPP STORE	<b>ANALISIS SENTIMEN MULTIVARIAT, AKURASI DAN PROTOKOL JARINGAN</b>
TOPIK 13	SISTEM BERBASI MAKASSAR APLIKASI TANAMAN MENGGUNAKAN PERANCANGAN PEMANFAATAN HIDROPONIK FAKTOR THING RANCANG OF INTERNET BANGUN ANALISI DIPA WEB UNIVERSITA MONITOR METOD MOBIL IOT PELAYANAN KENDARAAN IMPLEMENTASI KOTA DINA DARUSSALAM MERAH	<b>SISTEM BERBASIS IOT UNTUK PEMANTAUAN TANAMAN HIDROPONIK</b>

Tabel 1. Interpretasi Hasil LDA

Untuk menentukan celah penelitian, peneliti membandingkan hasil topik temuan dengan literatur terkait [10][5]. Dari literatur terkait masih ada topik yang mungkin belum tereksplorasi seperti potensi wisata, e-government, dan kecerdasan buatan (AI). Terdapat sejumlah pontensi penelitian baru dengan interseksi antara berbagai topik, misalnya memanfaatkan machine learning untuk memberikan rekomendasi destinasi wisata yang sesuai dengan preferensi individu, memanfaatkan data mining untuk meningkatkan efisiensi pelayanan publik dan pengambilan keputusan, pengembangan sistem rekomendasi mata kuliah berbasis AI yang mempertimbangkan minat, kemampuan, dan tujuan karier mahasiswa, memanfaatkan AI untuk mengembangkan sistem tutor cerdas yang dapat beradaptasi dengan gaya belajar setiap siswa, dll.

#### 4. Kesimpulan

Penelitian ini mengungkapkan 13 topik utama dalam judul-judul skripsi Teknik Informatika antara tahun 2022 hingga 2024 dengan menggunakan metode LDA, yang mencapai nilai koherensi 0,3. Hasil ini menunjukkan topik-topik yang dominan seperti data mining, machine learning, sistem berbasis web, android, analisis sentimen dan IoT, sekaligus mengidentifikasi beberapa area penelitian yang belum banyak dijelajahi tetapi memiliki potensi besar untuk dikembangkan. Nilai koherensi 0,3 menunjukkan bahwa meskipun ada keragaman topik yang dihasilkan, masih ada ruang untuk peningkatan dalam pemodelan topik yang lebih koheren. Rekomendasi dari penelitian ini adalah agar para peneliti mempertimbangkan topik-topik baru yang teridentifikasi untuk mendukung inovasi dan pengembangan ilmu pengetahuan di bidang Teknik Informatika. Peneliti merekomendasikan adanya kajian mendalam mengenai teknik pembersihan data untuk memastikan bahwa data yang digunakan dalam penelitian ini bebas dari noise dan istilah-istilah yang tidak relevan. Interpretasi bersifat subyektif, pengetahuan tentang domain yang spesifik sangat penting dalam menginterpretasi topik, terutama untuk data yang sangat khusus.

#### Daftar Pustaka

- [1] E. K. I. Nugraha, "Topic Modeling Dokumen Skripsi Prodi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta menggunakan Metode Latent Dirichlet Allocation," 2022.
- [2] Kamdan, Ivana Lucia Kharisma, Gina Purnama Insany, and Paikun, "Research topic modeling in informatics engineering study program at Nusa Putra University using LDA method," International



- Journal Engineering and Applied Technology (Ijeat), vol. 5, no. 2, pp. 24–35, 2022, doi: 10.52005/ijeat.v5i2.76.
- [3] M. I. Blei, D. M., Ng, A. Y., & Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, vol. 3 (jan), pp. 993–1022, 2003, doi: 10.1016/B978-0-12-411519-4.00006-9.
- [4] T. Intan Solihati and R. Kania, “Analisis Trend Dan Pemetaan Penelitian Mahasiswa Teknik Informatika Menggunakan Graph Convolutional Network,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 915–925, 2012, [Online]. Available: <http://jurnal.mdp.ac.id>
- [5] N. W. Utami and I. G. J. Eka Putra, “Text Minig Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine Similarity,” *Jurnal Informatika Teknologi dan Sains*, vol. 4, no. 3, pp. 255–259, 2022, doi: 10.51401/jinteks.v4i3.1907.
- [6] M. E. Moch Erreza, K. Kartini, and A. M. Rizki, “Pencarian Topik Penelitian Pada Studi Kasus Jurnal Jifti Menggunakan Teknik Hiearchical Dirichlet Processes,” *Spirit*, vol. 16, no. 1, pp. 170–182, 2024, doi: 10.53567/spirit.v16i1.325.
- [7] K. Jeong and Y. Kim, “Dynamic hierarchical Dirichlet processes topic model using the power prior approach,” *J Korean Stat Soc*, vol. 50, no. 3, pp. 860–873, 2021, doi: 10.1007/s42952-021-00129-1.
- [8] P. Kherwa and P. Bansal, “Topic Modeling: A Comprehensive Review,” *EAI Endorsed Transactions on Scalable Information Systems*, vol. 7, no. 24, pp. 1–16, 2019.
- [9] K. R. Nastiti, A. F. Hidayatullah, and A. R. Pratama, “Discovering Computer Science Research Topic Trends using Latent Dirichlet Allocation,” *Jurnal Online Informatika*, vol. 6, no. 1, p. 17, 2021, doi: 10.15575/join.v6i1.636.
- [10] N. Novarian, S. Khomsah, and A. B. Arifa, “Topic Modeling Tugas Akhir Mahasiswa Fakultas Informatika Institut Teknologi Telkom Purwokerto Menggunakan Metode Latent Dirichlet Allocation,” *LEDGER: Journal Informatic and Information Technology*, vol. 2, no. 1, pp. 14–27, 2023.
- [11] T. Titiana and D. H. Bangkalang, “Analisis Dan Penerapan Topic Modeling Pada Judul Tugas Akhir Mahasiswa Menggunakan Metode Latent Dirichlet Allocation (Lda),” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 4, pp. 1275–1287, 2023, doi: 10.29100/jipi.v8i4.4254.
- [12] C. Sievert and K. Shirley, “LDAvis: A method for visualizing and interpreting topics,” pp. 63–70, 2015, doi: 10.3115/v1/w14-3110.