

Topic Modelling Skripsi Manajemen Dakwah PTKIN menggunakan Latent Dirichlet Allocation (LDA)

Oryza Sativa Nufi, ²Khalid
Universitas Islam Negeri Sunan Ampel Surabaya^{1,2}

Surabaya, Indonesia

09030622051@uinsa.ac.id¹, khalid@uinsa.ac.id²

*Penulis Korespondensi

Diajukan : 25/03/2025

Diterima : 14/04/2025

Dipublikasi : 17/04/2025

ABSTRAK

Penulisan skripsi merupakan kewajiban akademik mahasiswa sebagai syarat kelulusan di perguruan tinggi. Program Studi Manajemen Dakwah di PTKIN se-Jawa memiliki keberagaman tema penelitian yang menarik untuk dianalisis lebih lanjut. Penelitian ini bertujuan untuk mengidentifikasi tren topik pada judul skripsi mahasiswa menggunakan metode pemodelan topik *Latent Dirichlet Allocation (LDA)*. Data yang digunakan berasal dari *repository digital* PTKIN, mencakup 1.648 judul skripsi dari tahun 2021 hingga 2023. Teknik pengambilan sampel dilakukan melalui *web scraping* dengan *library BeautifulSoup*, menghasilkan kumpulan judul yang dianalisis lebih lanjut. Hasil pemodelan LDA mengidentifikasi lima topik utama, yaitu pengelolaan lembaga dakwah, strategi komunikasi dakwah, peran masjid dan pesantren, serta pemanfaatan teknologi dalam dakwah. Validasi model menggunakan *coherence score* menunjukkan hasil 0,5107, yang mengindikasikan kualitas pemodelan yang baik. Hasil penelitian ini dapat menjadi referensi bagi mahasiswa dalam memilih tema skripsi yang relevan serta membantu institusi akademik dalam memahami perkembangan riset di bidang Manajemen Dakwah. Namun, keterbatasan dalam cakupan data menunjukkan perlunya studi lanjutan untuk memperluas analisis ke lebih banyak institusi pendidikan.

Kata Kunci: Latent Dirichlet Allocation, Manajemen Dakwah, Pemodelan Topik, Skripsi, Tren Penelitian

I. PENDAHULUAN

Penulisan skripsi merupakan salah satu kewajiban akademik yang harus diselesaikan oleh mahasiswa sebagai syarat kelulusan. Dalam konteks Program Studi Manajemen Dakwah di Perguruan Tinggi Keagamaan Islam Negeri (PTKIN), skripsi memiliki peran penting sebagai cerminan perkembangan keilmuan dan tren penelitian di program studi tersebut. Selain itu, skripsi juga menjadi bukti kemampuan mahasiswa dalam menganalisis dan memberikan solusi terhadap permasalahan dakwah. Manajemen Dakwah sebagai disiplin ilmu berfokus pada pengelolaan aktivitas dakwah, mencakup aspek strategi, komunikasi, hingga teknologi yang mendukung keberhasilan dakwah di tengah masyarakat.

Namun, meskipun Program Studi Manajemen Dakwah di Perguruan Tinggi Keagamaan Islam Negeri (PTKIN) memiliki konsentrasi keilmuan yang spesifik, mahasiswa sering kali menghadapi tantangan dalam menentukan tema penelitian yang relevan dengan kebutuhan keilmuan maupun praktik di masyarakat. Hal ini dapat disebabkan oleh kurangnya panduan atau pemetaan terhadap tema-tema penelitian yang telah dilakukan sebelumnya. Di PTKIN khususnya di Jawa, yang memiliki jumlah mahasiswa dan institusi yang cukup banyak, pemetaan tema penelitian menjadi semakin mendesak untuk menghindari duplikasi serta memastikan kontribusi

yang lebih signifikan bagi perkembangan ilmu Manajemen Dakwah. Tanpa pemahaman yang baik mengenai tren penelitian, terdapat risiko mahasiswa memilih tema yang kurang relevan atau bahkan mengulang topik yang sudah sering diteliti.

Seiring dengan perkembangan teknologi informasi, analisis terhadap tren penelitian kini dapat dilakukan dengan bantuan metode berbasis data, seperti *Latent Dirichlet Allocation (LDA)*. LDA adalah salah satu metode pemodelan topik yang mampu mengelompokkan dokumen teks ke dalam sejumlah topik tertentu berdasarkan pola distribusi kata-kata. Metode ini memungkinkan identifikasi tren penelitian dengan menganalisis kumpulan judul skripsi secara efisien. Selain LDA, metode lain dalam *topic modeling* seperti *Non-Negative Matrix Factorization (NMF)* dan *Latent Semantic Analysis (LSA)* juga dapat digunakan untuk mengidentifikasi pola dan kecenderungan penelitian. Dengan penerapan *topic modeling*, mahasiswa dan akademisi dapat memperoleh gambaran menyeluruh tentang tema-tema yang telah banyak diteliti serta menemukan celah penelitian baru yang belum banyak dieksplorasi.

Salah satu keunggulan utama LDA adalah kemampuannya dalam mengidentifikasi topik secara otomatis dari kumpulan dokumen besar tanpa memerlukan anotasi manual. Metode ini bekerja dengan mengasumsikan bahwa setiap dokumen terdiri dari beberapa topik, dan setiap topik memiliki sekumpulan kata yang sering muncul bersama. Keunggulan ini menjadikan LDA banyak digunakan dalam berbagai penelitian, termasuk analisis tren akademik, klasifikasi teks, serta ekstraksi informasi dari dokumen yang tidak terstruktur. Beberapa penelitian telah membuktikan efektivitas LDA dalam menganalisis tren penelitian skripsi di berbagai disiplin ilmu, seperti pendidikan, manajemen, dan studi keislaman. Dengan menerapkan LDA, analisis terhadap skripsi Manajemen Dakwah di PTKIN se-Jawa dapat dilakukan secara lebih sistematis, sehingga dapat memberikan wawasan yang lebih akurat mengenai arah penelitian yang berkembang di bidang ini.

Penelitian ini mengusulkan Topic Modelling skripsi Program Studi Manajemen Dakwah pada PTKIN se-Jawa menggunakan LDA dengan harapan dapat meningkatkan pemahaman terhadap pola dan kecenderungan penelitian di bidang ini. Dengan menganalisis judul-judul skripsi, metode ini memungkinkan identifikasi topik-topik utama yang muncul, sehingga dapat memberikan gambaran yang lebih jelas mengenai tren penelitian yang berkembang.

Beberapa penelitian terdahulu telah menerapkan metode Latent Dirichlet Allocation (LDA) untuk mengidentifikasi tren dalam berbagai konteks. (Kamdan et al., 2022) menggunakan LDA untuk menganalisis judul penelitian mahasiswa Teknik Informatika dan berhasil mengelompokkan tema dominan dari 159 judul. Sementara itu, (Alfanzar et al., 2020) menerapkan LDA dalam pemetaan topik skripsi mahasiswa Sastra Inggris di UIN Sunan Ampel Surabaya. Penelitian oleh (Xue et al., 2020) juga menunjukkan efektivitas LDA dalam mengekstraksi topik utama dari diskusi publik di media sosial selama pandemi COVID-19. Studi-studi ini memperlihatkan potensi LDA sebagai alat analisis teks yang mampu mengungkapkan pola tersembunyi dari data tidak terstruktur, termasuk judul skripsi. Namun, penerapannya dalam konteks Program Studi Manajemen Dakwah di PTKIN se-Jawa masih jarang dilakukan, sehingga penelitian ini penting untuk mengisi celah tersebut.

II. STUDI LITERATUR

Penelitian Terdahulu

Dari studi literatur yang dilakukan, didapatkan beberapa penggunaan LDA dalam berbagai studi kasus *topic modelling*, seperti Identifikasi Tren Penelitian Akademik, Analisis Diskursus Publik dan Sentimen Media Sosial, Klasifikasi Keluhan Konsumen.

LDA digunakan dalam beberapa *topic modeling* untuk Identifikasi Tren Penelitian Akademik. Penelitian sebelumnya menunjukkan bahwa metode Latent Dirichlet Allocation (LDA) dapat digunakan untuk mengidentifikasi tren penelitian mahasiswa di bidang Teknik Informatika di Universitas Nusa Putra. Dengan menerapkan LDA pada 159 judul penelitian, mereka berhasil mengelompokkan tema penelitian yang dominan (Kamdan et al., 2022). Pada penelitian lain, LDA juga digunakan dalam *topic modeling* skripsi sastra inggris di uin sunan ampel surabaya (Alfanzar et al., 2020). Studi ini memperkuat penggunaan LDA sebagai alat pemetaan tren akademik.

Pada studi kasus berikutnya, LDA diimplementasikan dalam Analisis Diskursus Publik dan

Sentimen Media Sosial. Penelitian oleh (Xue et al., 2020) menerapkan LDA untuk mengidentifikasi topik utama dalam diskusi publik selama pandemi COVID-19 di Twitter. Studi ini menemukan bahwa metode LDA efektif dalam mengekstraksi pola dari teks dalam jumlah besar, serta membantu mengelompokkan tema berdasarkan kemunculan kata-kata yang dominan. Meskipun penelitian ini tidak berfokus pada bidang akademik, hasilnya relevan dalam membuktikan kemampuan LDA dalam mengidentifikasi tren berdasarkan teks yang tidak terstruktur, seperti judul skripsi.

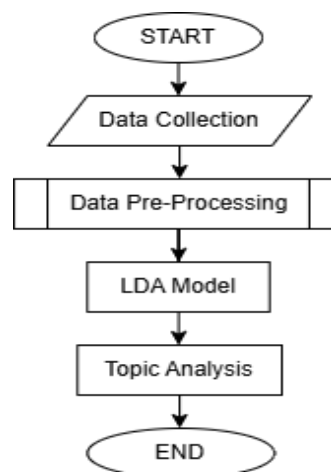
Penelitian lain oleh (Ahmed et al., 2022) juga menunjukkan bahwa LDA efektif dalam menganalisis tren penelitian dengan mengelompokkan topik berdasarkan teks yang tersedia. Dalam kajiannya terhadap berita ekonomi di Pakistan, metode ini mampu mengungkapkan pola utama dalam korpus yang besar dan membantu memahami kecenderungan tema dalam jangka waktu tertentu

Perihal perbandingan LDA dengan metode Pemodelan Topik Lain, Kajian oleh (Kherwa & Bansal, 2020) membandingkan LDA dengan metode pemodelan topik lainnya seperti *Latent Semantic Analysis (LSA)* dan *Non-Negative Matrix Factorization (NMF)*. Mereka menemukan bahwa LDA memiliki keunggulan dalam menghasilkan distribusi topik yang lebih akurat dan interpretatif dibandingkan metode lainnya. Studi ini memberikan dasar teoritis bahwa pemilihan LDA dalam penelitian skripsi Manajemen Dakwah di PTKIN merupakan keputusan yang tepat.

III. METODE

Penelitian ini menggunakan metode *Latent Dirichlet Allocation (LDA)* untuk menganalisis tema-tema yang terdapat pada judul skripsi mahasiswa Program Studi Manajemen Dakwah di PTKIN se-Jawa. Metode ini dipilih karena kemampuannya dalam mengidentifikasi topik-topik tersembunyi dari kumpulan dokumen teks yang besar secara probabilistik.

Menurut (Kamdan et al., 2022), LDA merupakan salah satu metode pemodelan topik yang populer karena kemampuannya dalam mengelompokkan dokumen berdasarkan distribusi kata-kata dalam teks. LDA telah banyak diterapkan dalam berbagai penelitian, termasuk untuk menganalisis tren penelitian akademik dengan menganalisis judul skripsi dan karya ilmiah mahasiswa. Berikut adalah tahapan penelitian yang dilakukan, dalam bentuk diagram alir:



Gambar 1. Flowchart Tahapan Penelitian

Data Collection

Data yang digunakan dalam penelitian ini berupa kumpulan judul skripsi mahasiswa Program Studi Manajemen Dakwah pada PTKIN se-Jawa untuk periode tahun 2021 hingga 2023. Pengumpulan data dilakukan melalui proses *crawling* dari *repository digital* PTKIN yang memiliki Program Studi Manajemen Dakwah. Proses *crawling* dilakukan dengan memanfaatkan skrip *Python* yang menggunakan *library BeautifulSoup* untuk mengakses, mengambil, dan menyimpan

data secara otomatis dari situs repositori masing-masing institusi.

Pemilihan PTKIN di Jawa sebagai objek penelitian didasarkan pada beberapa pertimbangan. Jawa merupakan wilayah dengan jumlah PTKIN terbanyak di Indonesia, sehingga dapat mewakili perkembangan keilmuan di bidang Manajemen Dakwah secara lebih komprehensif. Selain itu, PTKIN di Jawa memiliki jumlah mahasiswa yang besar serta akses terhadap teknologi dan sumber daya akademik yang lebih memadai dibandingkan wilayah lain. Dengan demikian, penelitian ini dapat memberikan gambaran yang lebih luas mengenai tren penelitian Manajemen Dakwah dan membantu mahasiswa serta institusi dalam menentukan arah penelitian di masa depan.

Data Pre-Processing

Sebelum melakukan pemodelan topik dengan metode *Latent Dirichlet Allocation (LDA)*, langkah pertama yang harus dilakukan adalah pra-pemrosesan data. Proses ini bertujuan untuk menyiapkan dan membersihkan data teks agar dapat dianalisis secara efektif. Data mentah yang diperoleh dari *crawling repository* PTKIN, yang berisi informasi seperti judul skripsi, abstrak, dan metadata lainnya, memerlukan beberapa tahap pengolahan untuk memastikan kualitas dan konsistensi data. Langkah-langkah *preprocessing* yang dilakukan meliputi *Case Folding*, *Tokenization*, *Stopword Removal*, *Stemming*, *Vectorization*.

LDA Modeling

Metode *Latent Dirichlet Allocation (LDA)* digunakan untuk melakukan analisis topik pada kumpulan judul dan abstrak skripsi. Membangun model *Latent Dirichlet Allocation (LDA)* dari awal berarti Anda membuat algoritma LDA sendiri tanpa menggunakan pustaka seperti *Gensim* atau *Scikit-learn*. Proses ini cukup kompleks karena melibatkan probabilitas, iterasi, dan distribusi statistik. Prosedur kerja LDA dapat dijelaskan sebagai berikut: (1) Inisialisasi beberapa parameter utama, termasuk jumlah dokumen, jumlah topik (k), dan jumlah iterasi. Dalam LDA, jumlah topik (k) merupakan parameter yang paling penting. (2) Menetapkan kata-kata ke dalam topik tertentu secara acak berdasarkan distribusi Dirichlet. (3) Melakukan iterasi pada setiap kata dalam korpus untuk memperbarui alokasi topik.

Topic Analysis

Analisis topik bertujuan untuk memahami dan menginterpretasikan hasil pemodelan LDA, mengidentifikasi tema-tema yang mendominasi dalam kumpulan data, serta menggali hubungan antar topik yang ditemukan. Pada tahap ini, kita akan menilai kualitas dan relevansi topik yang telah dihasilkan, serta mengaitkannya dengan konteks penelitian di bidang Manajemen Dakwah.

IV. HASIL DAN PEMBAHASAN

Pada bagian ini, dipaparkan hasil pemodelan topik menggunakan metode *Latent Dirichlet Allocation (LDA)* yang diterapkan pada judul-judul skripsi mahasiswa Program Studi Manajemen Dakwah di PTKIN se-Jawa. Hasil pemodelan menunjukkan beberapa topik dominan yang mencerminkan tren penelitian yang berkembang, yang dianalisis berdasarkan kata-kata kunci yang sering muncul. Pembahasan akan mengkaji relevansi topik-topik tersebut terhadap bidang Manajemen Dakwah serta memberikan wawasan tentang arah penelitian yang dapat dikembangkan di masa mendatang.

Hasil Pengumpulan Data

Dari proses pengambilan data berhasil dikumpulkan berjumlah 1.648 dataset abstrak dan judul skripsi Program Studi Manajemen Dakwah di PTKIN se-Jawa. Dengan rincian data setiap tahunnya dapat dilihat pada Tabel 2 berikut:

Tabel 1. Jumlah Dataset Abstrak Pertahun

Tahun	Jumlah Abstrak
-------	----------------

2021	489
2022	587
2023	572

Data PreProcessing

Sebelum melakukan pemodelan topik dengan metode *Latent Dirichlet Allocation (LDA)*, langkah pertama yang harus dilakukan adalah pra-pemrosesan data. Proses ini bertujuan untuk menyiapkan dan membersihkan data teks agar dapat dianalisis secara efektif. Kami akan melakukan *preprocessing* pada kolom “Judul” dan “Abstrak” sesuai dengan *Data Collection* sebelumnya.

Tabel 2. Hasil Data PreProcessing

	Sebelum Preprocessing	Setelah Preprocessing
Judul	Manajemen program kerja seksi dakwah dan ibadah masjid Al-Ikhlas Sidosermo Surabaya.	[manajemen, program, kerja, seksi, dakwah, ibadah, masjid, ikhlas, sidosermo, surabaya]
Abstrak	Penelitian ini terfokus pada rumusan masalah bagaimana program kerja seksi dakwah dan ibadah masjid Al-Ikhlas Sidosermo Surabaya? Untuk menjawab rumusan masalah tersebut penulis menggunakan metode penelitian deskriptif kualitatif. Informan dalam penelitian ini adalah ketua takmir, seksi dakwah dan ibadah, dan pengurus harian. Data dikumpulkan dengan teknik wawancara, observasi, dan dokumentasi. Data yang terkumpul dianalisis dianalisis dengan menggunakan teknik analisis interkatif Miles dan Huberman yang meliputi tiga komponen: reduksi data, penyajian data, dan penarikan serta pengujian kesimpulan. Dalam penelitian ini penulis berusaha untuk mendeskripsikan manajemen program kerja seksi dakwah dan ibadah masjid Al-Ikhlas Sidosermo. Penulis memanfaatkan beberapa teori yang terkait dengan permasalahan tersebut. Hasil penelitian menyimpulkan bahwa pelaksanaan manajemen program kerja seksi dakwah dan ibadah Masjid Al-Ikhlas Sidosermo Surabaya berjalan secara baik melalui kegiatan perencanaan, pengorganisasian, penggerakan, dan pengawasan.	[penelitian, fokus, rumus, masalah, program, kerja, seksi, dakwah, ibadah, masjid, ikhlas, sidosermo, surabaya, jawab, rumus, masalah, penulis, guna, metode, penelitian, deskriptif, kualitatif, informan, penelitian, ketua, takmir, seksi, dakwah, ibadah, pengurus, harian, data, kumpul, teknik, wawancara, observasi, dokumentasi, data, kumpul, analisis, analisis, guna, teknik, analisis, interaktif, miles, huberman, liput, tiga, komponen, reduksi, data, sajian, data, tarik, uji, kesimpulan, penelitian, penulis, usaha, deskripsi, manajemen, program, kerja, seksi, dakwah, ibadah, masjid, ikhlas, sidosermo, penulis, manfaat, teori, kait, masalah, hasil, penelitian, simpul, pelaksana, manajemen, program, kerja, seksi, dakwah, ibadah, masjid, ikhlas, sidosermo, surabaya, jalan, baik, kegiatan, rencana, organisasi, gerak, awas]

LDA Topic Modeling

Setelah tahap pra-pemrosesan data selesai, metode *Latent Dirichlet Allocation (LDA)* digunakan untuk mengidentifikasi topik-topik tersembunyi dalam judul-judul skripsi yang telah dikumpulkan. LDA bekerja dengan membagi kumpulan dokumen menjadi beberapa topik berdasarkan distribusi kata-kata yang sering muncul bersama. Dalam penelitian ini, model LDA dibuat dengan jumlah topik yang telah ditentukan, menggunakan korpus data hasil pra-pemrosesan. Hasil dari pemodelan ini berupa kumpulan topik yang diwakili oleh kata-kata kunci dominan, yang menggambarkan tema utama dari setiap topik.

Vectorization adalah langkah pertama dalam mempersiapkan input untuk algoritma LDA. Pada bagian ini, teks *tokenized* diubah menjadi representasi numerik, yaitu *Bag of Words (BoW)*, menggunakan proses *vectorization*.

Tabel 3. Data setelah menghasilkan Fitur

Fitur (Kata)
['adha', 'alquran', 'analisis', 'awas', 'baik', 'bulan', 'cukup', 'dakwah', 'dana', 'data', 'daya', 'deskripsi', 'deskriptif', 'dokumentasi', 'dukung', 'erat', 'error', 'faktor', 'fardlu', 'fitri', 'fokus', 'gerak', 'gerhana', 'guna', 'hajat', 'hambat', 'harian', 'hasil', 'huberman', 'human', 'ibadah', 'idul', 'ikhlas', 'infaq', 'informan', 'insiden', 'interaktif', 'isra', 'isyraq', 'jalan', 'jawab', 'jumat', 'kait', 'kegiatan', 'kekeluargaan', 'kelompok', 'kerja', 'kesimpulan', 'ketua', 'komponen', 'kualitatif', 'kumpul', 'kurang', 'lima', 'liput', 'manajemen', 'manfaat', 'manusia', 'masalah', 'masjid', 'masyarakat', 'maulid', 'metode', 'mi', 'miles', 'minggu', 'nabi', 'observasi', 'organisasi', 'pelaksana', 'penelitian', 'pengurus', 'penulis', 'peringatan', 'program', 'qurban', 'raj', 'rasa', 'reduksi', 'rencana', 'respons', 'rumus', 'rutin', 'sadar', 'sajian', 'seksi', 'shadaqah', 'shalat', 'sidosermo', 'sifat', 'simpul', 'sumber', 'surabaya', 'takmir', 'taraweh', 'tarik', 'tasbih', 'taubat', 'teknik', 'temu', 'teori', 'tiga', 'tilawah', 'uji', 'usaha', 'waktu', 'wawancara', 'zakat']

Tabel 4. Data setelah Vectorization

	adha	alquran	analisis	awas	baik	...	waktu	wawancara
Judul	0	0	0	0	0	...	0	0
Abstrak	1	1	3	1	2	...	1	1

Matriks vektor yang dihasilkan dari proses *vectorization* menggunakan metode *Bag of Words (BoW)* merepresentasikan dokumen-dokumen dalam bentuk vektor numerik yang dapat diproses lebih lanjut dalam model analisis teks. Dalam konteks penelitian ini, matriks vektor digunakan untuk menggambarkan teks dalam judul dan abstrak sebagai sekumpulan angka yang merefleksikan keberadaan atau frekuensi kata-kata yang terkandung di dalamnya.

Setelah model *Latent Dirichlet Allocation (LDA)* dilatih, kualitas topik dievaluasi menggunakan metrik *coherence score*. Metrik ini mengukur sejauh mana kata-kata dalam topik memiliki keterkaitan semantik yang kuat, dengan nilai koherensi yang lebih tinggi menunjukkan topik yang lebih bermakna dan relevan. Pada penelitian ini, nilai koherensi dihitung untuk setiap topik menggunakan *corpus* dan *dictionary* sebagai referensi, untuk memastikan bahwa jumlah topik yang dipilih sudah optimal. Hasil nilai koherensi disajikan untuk memberikan gambaran kualitas topik yang dihasilkan oleh model.

Tabel 5. Hasil Nilai Koherensi Pada Setiap Topik

Topik	Kata
Topik 1	0.053*"masjid" + 0.052*"dakwah" + 0.020*"manajemen" + 0.016*"masyarakat" + 0.015*"islam" + 0.015*"agama" + 0.011*"organisasi" + 0.010*"jamaah" + 0.009*"rencana" + 0.009*"semarang"
Topik 2	0.040*"kerja" + 0.032*"haji" + 0.029*"layan" + 0.020*"agama" + 0.019*"nilai" + 0.015*"menteri" + 0.015*"kantor" + 0.014*"pegawai" + 0.011*"kabupaten" + 0.011*"umrah"
Topik 3	0.049*"pesantren" + 0.043*"pondok" + 0.033*"santri" + 0.018*"manajemen" + 0.016*"pimpin" + 0.015*"organisasi" + 0.014*"kembang" + 0.013*"program" + 0.012*"sumber" + 0.011*"kualitas"
Topik 4	0.027*"wisata" + 0.020*"makam" + 0.014*"religi" + 0.014*"kelola" + 0.013*"kembang" + 0.012*"manajemen" + 0.012*"kabupaten" + 0.011*"desa" + 0.008*"masyarakat" + 0.007*"daya"
Topik 5	0.041*"zakat" + 0.019*"program" + 0.015*"dana" + 0.014*"lembaga" + 0.013*"baznas" + 0.011*"kelola" + 0.010*"manajemen" + 0.010*"kabupaten" + 0.009*"masyarakat" + 0.009*"kota"

Berdasarkan hasil *processing* yang kami lakukan, telah dihasilkan *Coherence Score* sebesar 0.5107 menunjukkan tingkat koherensi topik yang dihasilkan oleh model *Latent Dirichlet Allocation (LDA)*. Dalam konteks analisis topik, *Coherence Score* digunakan untuk mengukur sejauh mana kata-kata dalam setiap topik memiliki keterkaitan yang bermakna.

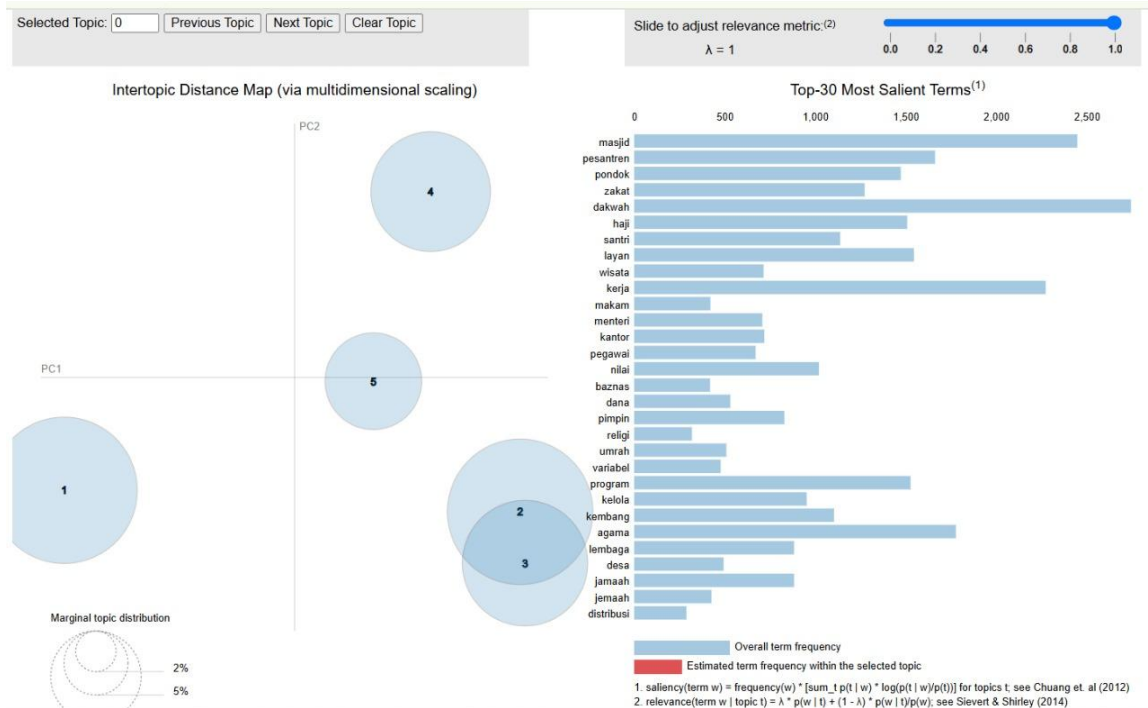
Topic Analysis

Setelah pemodelan *Latent Dirichlet Allocation (LDA)* diterapkan, tahap analisis dilakukan untuk memahami distribusi topik yang dihasilkan, mengevaluasi kualitas model, serta mengidentifikasi tren penelitian dalam skripsi mahasiswa Program Studi Manajemen Dakwah di PTKIN se-Jawa.

Hasil pemodelan menunjukkan adanya lima topik utama, yang diinterpretasikan berdasarkan kata-kata kunci yang paling dominan dalam setiap topik. Beberapa tema yang ditemukan meliputi pengelolaan lembaga dakwah, strategi komunikasi dakwah, peran masjid dan pesantren, serta pemanfaatan teknologi dalam dakwah. Kata-kata kunci yang sering muncul dalam setiap topik dianalisis untuk menentukan relevansi dan keterkaitannya dengan bidang Manajemen Dakwah.

Untuk memastikan kualitas pemodelan, dilakukan evaluasi menggunakan *coherence score*, yang dalam penelitian ini memiliki nilai 0,5107. Nilai ini menunjukkan bahwa pemodelan LDA berhasil menghasilkan topik yang memiliki keterkaitan semantik yang cukup baik.

Untuk mempermudah interpretasi hasil pemodelan LDA, kami melakukan visualisasi menggunakan *library pyLDAvis*, yang membantu memvisualisasikan hubungan antara topik-topik serta distribusi kata-kata di setiap topik. Visualisasi ini memberikan wawasan yang lebih intuitif tentang bagaimana topik-topik terbentuk dan relevansi setiap kata dalam topik tertentu.



Gambar 2. Visualisasi menggunakan pyLDAvis

Berdasarkan hasil visualisasi tersebut, topik-topik yang dihasilkan oleh model *Latent Dirichlet Allocation (LDA)* dapat dianalisis lebih mendalam melalui dua elemen utama:

1. Intertopic Distance Map

Pada peta sebelah kiri, lingkaran-lingkaran menunjukkan distribusi topik yang dihasilkan oleh model LDA. Posisi lingkaran merepresentasikan jarak antar topik dalam ruang multidimensi, sementara ukuran lingkaran menunjukkan proporsi dari setiap topik dalam dataset. Dari visualisasi ini, terlihat bahwa terdapat lima topik utama (Topik 1 hingga Topik 5), dengan beberapa topik yang saling berdekatan (menandakan kemiripan) dan lainnya yang lebih terpisah (menunjukkan perbedaan yang signifikan).

2. Top-30 Most Salient Terms

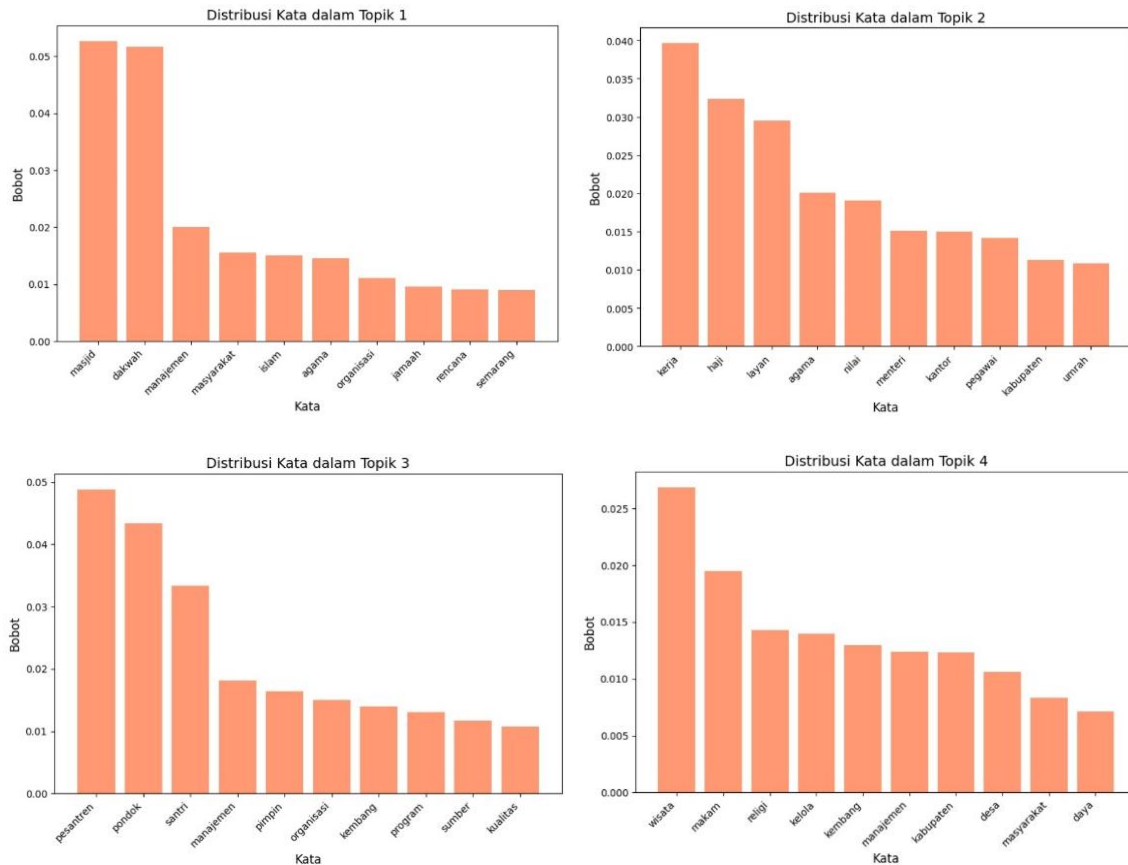
Grafik batang di sebelah kanan menunjukkan 30 istilah paling penting yang ditemukan dalam seluruh topik. Kata-kata seperti "masjid", "pesantren", "pondok", "zakat", dan "dakwah" memiliki frekuensi yang tinggi dan menjadi ciri utama dari topik yang diidentifikasi. Warna merah pada grafik menunjukkan frekuensi istilah tersebut di topik tertentu, sementara warna biru menggambarkan frekuensi keseluruhannya di *corpus*.

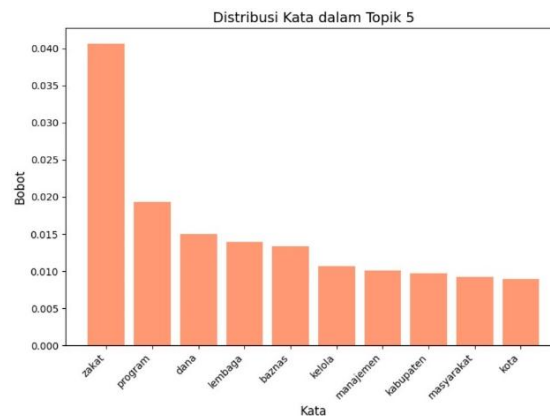




Gambar 3. Visualisasi dengan Word Cloud

Visualisasi *word cloud* dalam penelitian ini dibuat berdasarkan hasil pemodelan LDA, sebagaimana ditunjukkan dalam Tabel 5. Dengan demikian, *word cloud* memvisualisasikan lima topik utama yang dihasilkan dari proses LDA. Setiap *word cloud* menggambarkan kata-kata yang paling relevan dengan masing-masing topik. Kata-kata yang lebih besar menunjukkan frekuensi atau relevansi yang lebih tinggi dalam topik tersebut.





Gambar 4. Visualisasi Jumlah Kata (Word Count)

Visualisasi jumlah kata (*word count*) adalah representasi grafis yang menunjukkan frekuensi kemunculan setiap kata dalam suatu kumpulan teks. Dalam konteks penelitian ini, visualisasi *word count* memberikan gambaran mengenai distribusi dan dominasi kata-kata yang sering muncul dalam teks yang telah diproses. Visualisasi ini berfungsi untuk menganalisis kata-kata kunci yang paling sering digunakan dan memberikan wawasan tentang tema-tema utama dalam teks yang dianalisis.

V. KESIMPULAN

Berdasarkan hasil penelitian ini, dalam kurun waktu tiga tahun (2021–2023), diperoleh total 1.648 judul skripsi dari Program Studi Manajemen Dakwah di PTKIN se-Jawa. Dengan menerapkan metode *Latent Dirichlet Allocation (LDA)*, penelitian ini mengidentifikasi lima topik utama yang mencerminkan tren penelitian di bidang Manajemen Dakwah. Topik-topik yang dominan meliputi pengelolaan lembaga dakwah, strategi komunikasi dakwah, peran masjid dan pesantren, serta pemanfaatan teknologi dalam dakwah. Validasi model menggunakan *coherence score* menunjukkan bahwa pemodelan topik memiliki kualitas yang baik dengan tingkat koherensi sebesar 0.5107. Dalam analisis tren topik, ditemukan bahwa topik terkait masjid dan pesantren memiliki frekuensi kemunculan tertinggi dalam penelitian mahasiswa. Visualisasi menggunakan *pyLDAvis* dan *word cloud* memperjelas distribusi dan keterkaitan antar topik. Hasil penelitian ini dapat menjadi referensi bagi mahasiswa dalam menentukan topik skripsi yang relevan serta membantu institusi akademik dalam memahami perkembangan riset di bidang Manajemen Dakwah. Namun, keterbatasan dalam cakupan data menunjukkan perlunya studi lanjutan untuk memperluas analisis ke lebih banyak institusi pendidikan.

VI. REFERENSI

- Ahmed, F., Nawaz, M., & Jadoon, A. (2022). Topic Modeling of the Pakistani Economy in English Newspapers via Latent Dirichlet Allocation (LDA). *SAGE Open*, 12(1). <https://doi.org/10.1177/21582440221079931>
- Alfanzar, A. I., Khalid, K., & Rozas, I. S. (2020). Topic Modelling Skripsi Menggunakan Metode Latent Dirichlet Allocation. *JSiI (Jurnal Sistem Informasi)*, 7(1), 7. <https://doi.org/10.30656/jsii.v7i1.2036>
- Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Systems with Applications*, 127, 256–271. <https://doi.org/10.1016/j.eswa.2019.03.001>
- Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*, 163, 1–13. <https://doi.org/10.1016/j.knosys.2018.08.011>

- Du, Y. J., Yi, Y. T., Li, X. Y., Chen, X. L., Fan, Y. Q., & Su, F. H. (2020). Extracting and tracking hot topics of micro-blogs based on improved Latent Dirichlet Allocation. *Engineering Applications of Artificial Intelligence*, 87(September 2019), 103279. <https://doi.org/10.1016/j.engappai.2019.103279>
- Hakim, R., Khalid, K., & Rolliawati, D. (2022). Topic Modeling Pada Abstrak Skripsi Menggunakan Metode Latent Semantic Analysis. *Format : Jurnal Ilmiah Teknik Informatika*, 11(1), 83. <https://doi.org/10.22441/10.22441/format.2022.v11.i1.009>
- Heo, S., & Yang, J. (2020). *Jksoci 1*. 25(12), 83–91.
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>
- Kamdan, Ivana Lucia Kharisma, Gina Purnama Insany, & Paikun. (2022). Research topic modeling in informatics engineering study program at Nusa Putra University using LDA method. *International Journal Engineering and Applied Technology (Ijeat)*, 5(2), 24–35. <https://doi.org/10.52005/ijeat.v5i2.76>
- Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, 7(24), 1–16. <https://doi.org/10.4108/eai.13-7-2018.159623>
- Learning, D. (2022). *View of Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation | The International Review of Research in Open and Distributed Learning*. <http://www.irrodl.org/index.php/irrodl/article/view/5358/5506>
- Lee, J. H., & Ostwald, M. J. (2024). Latent Dirichlet Allocation (LDA) topic models for Space Syntax studies on spatial experience. *City, Territory and Architecture*, 11(1). <https://doi.org/10.1186/s40410-023-00223-3>
- Min, K. B., Song, S. H., & Min, J. Y. (2020). Topic modeling of social networking service data on occupational accidents in Korea: Latent dirichlet allocation analysis. *Journal of Medical Internet Research*, 22(8), 1–12. <https://doi.org/10.2196/19222>
- Nawang Sari, W. A., & Dwi Purnomo, H. (2022). Topic Modeling Using the Latent Dirichlet Allocation Method on Wikipedia Pandemic Covid-19 Data in Indonesia. *Jurnal Teknik Informatika (Jutif)*, 3(5), 1223–1230. <https://doi.org/10.20884/1.jutif.2022.3.5.321>
- Negara, E. S., & Triadi, D. (2022). Topic modeling using latent dirichlet allocation (LDA) on twitter data with Indonesia keyword. *Bulletin of Social Informatics Theory and Application*, 5(2), 124–132. <https://doi.org/10.31763/businta.v5i2.455>
- Sharma, C., Batra, I., Sharma, S., Malik, A., Sanwar Hosen, A. S. M., & Ra, I. H. (2022). Predicting Trends and Research Patterns of Smart Cities: A Semi-Automatic Review Using Latent Dirichlet Allocation (LDA). *IEEE Access*, 10(November), 121080–121095. <https://doi.org/10.1109/ACCESS.2022.3214310>
- Sharma, C., Sharma, S., & Sakshi. (2022). Latent DIRICHLET allocation (LDA) based information modelling on BLOCKCHAIN technology: a review of trends and research patterns used in integration. *Multimedia Tools and Applications*, 81(25), 36805–36831. <https://doi.org/10.1007/s11042-022-13500-z>
- Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter. *PLoS ONE*, 15(9 September), 1–12. <https://doi.org/10.1371/journal.pone.0239441>
- Zhou, S., Kan, P., Huang, Q., & Silbernagel, J. (2023). A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura. *Journal of Information Science*, 49(2), 465–479. <https://doi.org/10.1177/01655515211007724>