



**ANALISIS KOMPARATIF BERTOPIC DAN LDA UNTUK
PEMODELAN TOPIK PUBLIKASI AKADEMIK PADA
DATASET GOOGLE SCHOLAR**

**TUGAS AKHIR
SKRIPSI**

RIZQI ALIF PERMANA PUTRA

41522010001

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA**

2026



**ANALISIS KOMPARATIF BERTOPIC DAN LDA UNTUK
PEMODELAN TOPIK PUBLIKASI AKADEMIK PADA
DATASET GOOGLE SCHOLAR**

**TUGAS AKHIR
SKRIPSI**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

RIZQI ALIF PERMANA PUTRA
41522010001

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Rizqi Alif Permana Putra

NIM : 41522010001

Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“Analisis Komparatif BERTopic dan LDA Untuk Pemodelan Topik Publikasi Akademik Pada Dataset Google Scholar” adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta, atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau hak pihak manapun.

Apabila di kemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

UNIVERSITAS Jakarta, 28 Juli 2026

MERCU BUANA



Rizqi Alif Permana Putra.

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name : Rizqi Alif Permana Putra
NIM /Student ID Number : 41522010001
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

**“ANALISIS KOMPARATIF BERTOPIC DAN LDA UNTUK PEMODELAN TOPIK
PUBLIKASI AKADEMIK PADA DATASET GOOGLE SCHOLAR”**

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

18 Juli 2026

dengan nilai persentase sebesar :

/with a percentage value of:

5%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer**
Universitas Mercu Buana. */declared to meet standards in accordance with applicable
regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

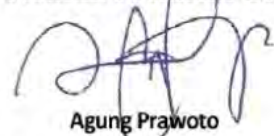
File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1p3j8XSUkM8L5KNgrlFEP46rxWpcFimFL/view?usp=drive_link



Jakarta, 18 Juli 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Rizqi Alif Permana Putra
NIM : 41522010001
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : Analisis Komparatif BERTopic dan LDA Untuk
Pemodelan Topik Publikasi Akademik Pada Dataset
Google Scholar

Telah berhasil dipertahankan pada sidang tanggal 28 Juli 2026 dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing

Dr. Ir. Afiyati, S.Si., M.T

NIDN/NUPTK: 0316106908

UNIVERSITAS

Jakarta, 28 Juli 2026

Mengetahui,

MERCU BUANA

Dekan Fakultas Ilmu Komputer

Dr. Bambang Jokonowo, S.Si., MTI
NIDN/NUPTK: 0320037002

Ketua Program Studi
Teknik Informatika

Dr. Hadi Santoso, S.Kom., M.Kom
NIDN/NUPTK: 0225067701

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Tugas Akhir ini. Penulisan Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Teknik Informatika pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Tugas Akhir ini, sangatlah sulit bagi saya untuk menyelesaikan Tugas Akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer Universitas Mercu Buana.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
4. Bapak/Ibu Dr. Ir. Afiyati S.Si., M.T. selaku dosen pembimbing tugas akhir yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan tugas akhir ini terjadwal dengan baik.
5. Kedua Orang Tua saya yang senantiasa memberikan doa, kasih sayang, dukungan moral maupun material, serta semangat selama penulis menempuh pendidikan di Universitas Mercu Buana.
6. Dita Istiana Utami, yang telah memberikan dukungan, motivasi, perhatian, serta semangat kepada penulis sehingga dapat tetap berusaha dan menyelesaikan tugas akhir ini dengan baik.
7. Seluruh teman-teman kuliah, khususnya rekan-rekan Teknik Informatika Universitas Mercu Buana, yang telah menjadi teman berdiskusi, saling membantu, berbagi pengalaman, dan memberikan dukungan selama kurang lebih empat tahun menjalani masa perkuliahan hingga penyusunan tugas akhir ini.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Tugas Akhir ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 28 Juli 2026

Rizqi Alif Permana Putra



**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR DI REPOSITORI UMB**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Rizqi Alif Permaan Putra
NIM : 41522010001
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : Analisis Komparatif BERTopic dan LDA
Untuk Pemodelan Topik Publikasi Akademik
Pada Dataset Google Scholar

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 28 Juli 2026

Yang menyatakan,



Rizqi Alif Permana Putra

ANALISIS KOMPARATIF BERTOPIC DAN LDA UNTUK PEMODELAN TOPIK PUBLIKASI AKADEMIK PADA DATASET GOOGLE SCHOLAR

RIZQI ALIF PERMANA PUTRA

ABSTRAK

Pemetaan kepakaran tenaga pendidik pada pangkalan data akademik dilaksanakan untuk meminimalkan redundansi arah penelitian melalui pendekatan penambangan teks terstruktur. Ekstraksi struktur tema laten dalam penelitian ini diimplementasikan melalui komparasi arsitektur pemodelan topik Latent Dirichlet Allocation (LDA) berbasis probabilitas statistik dan BERTopic berbasis semantic embedding. Korpus dokumen dibentuk melalui integrasi atribut judul publikasi dan bidang keilmuan dari repositori Google Scholar, yang menghasilkan 1.879 dokumen tervalidasi pada rentang tahun 2015 hingga 2026. Evaluasi kuantitatif mengalkulasi perolehan Coherence Score (Cv) sebesar 0,5804 untuk arsitektur LDA pada konfigurasi 5 topik laten, sedangkan arsitektur BERTopic memproyeksikan skor 0,5655 pada konfigurasi 12 klaster semantik. Analisis komparatif membuktikan bahwa algoritma LDA memiliki kinerja komputasi yang lebih optimal dalam mengidentifikasi pemusatan orientasi riset institusi pada domain Data Science dan Artificial Intelligence. Temuan ini mengonfirmasi bahwa kedua model mampu memetakan dinamika kepakaran tenaga pendidik secara terstruktur, dengan LDA menunjukkan keunggulan stabilitas koherensi pada dataset akademik yang terbatas. Hasil penelitian ini direkomendasikan sebagai instrumen audit kuantitatif bagi institusi dalam mengevaluasi produktivitas riset serta merumuskan peta jalan penelitian berbasis data aktual.

Kata kunci: Pemodelan Topik, Latent Dirichlet Allocation, BERTopic, Google Scholar, Coherence Score.

ANALISIS KOMPARATIF BERTOPIC DAN LDA UNTUK PEMODELAN TOPIK PUBLIKASI AKADEMIK PADA DATASET GOOGLE SCHOLAR

RIZQI ALIF PERMANA PUTRA

ABSTRACT

The mapping of educators' expertise within academic databases is conducted to minimize research redundancy through a structured text mining approach. This research implements the extraction of latent theme structures by comparing two topic modeling architectures: Latent Dirichlet Allocation (LDA) based on a Bag-of-Words matrix and BERTopic based on spatial dimension reduction and semantic embeddings. The document corpus was constructed by integrating publication titles and scientific field attributes from the Google Scholar repository, resulting in 1,879 validated documents published between 2015 and 2026. Quantitative evaluation using the Coherence Score (C_v) yielded a score of 0.5804 for the LDA architecture with a 5-topic configuration, while the BERTopic architecture projected a score of 0.5655 with a 12-cluster configuration. Comparative analysis demonstrates that the LDA algorithm offers superior computational performance in identifying institutional research orientation, specifically within the domains of Data Science and Artificial Intelligence, evidenced by an accumulation of 267 academic publications. Consequently, the LDA architecture is established as the optimal analytical model for visualizing research roadmaps and expertise trend dynamics within educational institutions.

Kata kunci: Topic Modeling, Latent Dirichlet Allocation, BERTopic, Google Scholar, Coherence Score.

DAFTAR ISI

HALAMAN SAMPUL.....	0
HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB.....	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	3
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 Teori Pendukung	15
2.2.1 Pemodelan Topik	15
2.2.2 Latent Dirichlet Allocation (LDA)	15
2.2.3 BERTopic	16
2.2.4 Metrik Evaluasi Model.....	17
2.3 Gap Penelitian	17
BAB III METODE PENELITIAN	20
3.1 Jenis Penelitian.....	20
3.2 Tahapan Penelitian	20

BAB IV PEMBAHASAN.....	26
4.1 Pengumpulan Data	26
4.2 Dataset.....	26
4.3 Analisis Data	28
4.3.1 Import Library	28
4.3.2 Validasi & Pembersihan Data	31
4.4 Pra-Processing.....	33
4.4.1 Inisialisasi Stemmer.....	33
4.4.2 Penyusunan Daftar Stopword	34
4.4.3 Pembangunan Pipeline Pembersihan Teks	34
4.4.4 Transformasi dan Penyimpanan Dataset.....	35
4.4.5 Tokenisasi dan Pembangunan Kamus Frekuensi	36
4.5 Transformasi Data.....	38
4.5.1 Pemodelan Algoritma Latent Dirichlet Allocation (LDA).....	38
4.5.2 Pemodelan Algoritma BERTopic.....	39
4.6 Evaluasi Model	40
4.6.1 Hasil Topic Modeling LDA.....	41
4.6.2 Hasil Topic Modeling LDA.....	42
4.6.3 Hasil Perhitungan Coherence Score	45
4.7 Hasil Visualisasi Topic	46
4.7.1 Grafik Koherensi.....	46
4.7.2 WordCloud.....	49
4.7.3 Diagram Garis	52
4.8 Topik Dominan	56
BAB V KESIMPULAN DAN SARAN	58
5.1 Kesimpulan	58
5.2 Saran.....	59
DAFTAR PUSTAKA.....	61
LAMPIRAN.....	64

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	6
Tabel 2. 2 Gap Penelitian	18
Tabel 4. 1 Sampel Dataset.....	26
Tabel 4. 2 Struktur Dataframe.....	32
Tabel 4. 3 Pra-Processing Data	37
Tabel 4. 4 Hasil Topic Modeling dengan LDA-BoW.....	41
Tabel 4. 5 Hasil Topic Modeling BERTopic	43
Tabel 4. 6 Hasil Perhitungan Coherence Score.....	45



DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian	21
Gambar 4. 1 Import Library	29
Gambar 4. 2 Validasi Data	31
Gambar 4. 3 Stemmer	33
Gambar 4. 4 Penyusunan Daftar Stopword	34
Gambar 4. 5 Pipeline Pembersihan Teks	35
Gambar 4. 6 Transformasi Dataset	36
Gambar 4. 7 Tokenisasi dan Pembangunan Kamus	36
Gambar 4. 8 Model Algoritma LDA	38
Gambar 4. 9 Model BERTopic	40
Gambar 4. 10 Grafik Koherensi LDA	47
Gambar 4. 11 Grafik Koherensi BERTopic	48
Gambar 4. 12 WordCloud LDA	50
Gambar 4. 13 WordCloud BERTopic	51
Gambar 4. 14 Dinamika Tren Topik LDA	53
Gambar 4. 15 Dinamika Tren Topik BERTopic	55



DAFTAR LAMPIRAN

Lampiran 1. Kartu Asistensi	64
Lampiran 2. Curriculum Vitae	65
Lampiran 3. Surat Pernyataan & Pengalihan HAKI	66
Lampiran 4. Surat Keterangan BNSP	68
Lampiran 5. Mapping Nama Dosen	69

