



**PEMODELAN TOPIK KONTEKSTUAL TEKS HADIS
KATEGORI PERINTAH, LARANGAN, DAN INFORMASI
MENGUNAKAN BERTOPIC BERBASIS INDOBERT**

**TUGAS AKHIR
SKRIPSI**

RENDY ARIESTA CHANDRA

41522010139

**UNIVERSITAS
MERCU BUANA**

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



**PEMODELAN TOPIK KONTEKSTUAL TEKS HADIS
KATEGORI PERINTAH, LARANGAN, DAN INFORMASI
MENGUNAKAN BERTOPIC BERBASIS INDOBERT**

**TUGAS AKHIR
SKRIPSI**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

RENDY ARIESTA CHANDRA

41522010139

**UNIVERSITAS
MERCU BUANA**

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

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Rendy Ariesta Chandra
NIM : 41522010139
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“Pemodelan Topik Kontekstual Teks Hadis Kategori Perintah, Larangan, Dan Informasi Menggunakan Bertopic Berbasis Indobert” 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
MERCU BUANA

Jakarta, 05 Agustus 2026



Rendy Ariesta Chandra

HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN



072.423.4.03.01

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 : Rendy Ariesta Chandra
NIM /Student ID Number : 41522010139
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir
(The title:

“PEMODELAN TOPIK KONTEKSTUAL TEKS HADIS KATEGORI PERINTAH, LARANGAN, DAN INFORMASI MENGGUNAKAN BERTOPIC BERBASIS INDOBERT”

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:
(Similarity checks have been carried out with the Turnitin system on the date:

1 Juli 2026

dengan nilai persentase sebesar :
(with a percentage value of:

6%

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/1S1WGKR71XXRV9yGiNuziNZ81J05wVR0P/view?usp=drive_link



Jakarta, 1 Juli 2026
Admin Turnitin Fasilkom UMB

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

Fakultas Ilmu Komputer
KAMPUS MENARA BHAKTI
Jl. Raya Meruya Selatan No. 1 Kembangan, Jakarta Barat 11650
Telp. 021-5840816 (Hunting), Psw : 5700 Fax. 021-5840813
<http://www.mercubuana.ac.id>, e-mail : fasilkom@mercubuana.ac.id

HALAMAN PENGESAHAN

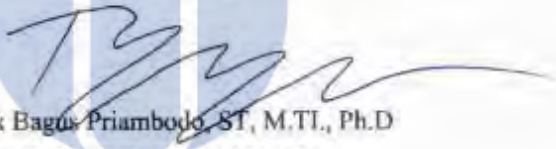
Tugas Akhir ini diajukan oleh:

Nama : RENDY ARIESTA CHANDRA
NIM : 41522010139
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : Topik Kontekstual Teks Hadis Kategori Perintah,
Larangan, dan Informasi Menggunakan BERTopic
Berbasis IndoBERT.

Telah berhasil dipertahankan pada sidang tanggal 25 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


Bapak Bagus Priambodo, ST, M.TI., Ph.D

NIDN/NUPTK: 0313057905

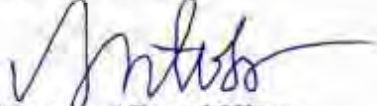
Jakarta, 05 Agustus 2026

Mengetahui,

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.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
4. Bapak Dr. Bagus Priambodo, S.T., M.TI. 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. Bapak Suharyadi selaku Kepala Sub. Jaringan yang telah memberikan dukungan penuh, termasuk dalam fasilitas penunjang, sehingga saya dapat menyelesaikan laporan Tugas Akhir ini dengan baik.
6. Kedua Orang Tua saya yang selalu mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercu Buana.
7. Semua teman kuliah yang selalu berbagi informasi dan memberikan dukungan dalam bentuk yang berbeda-beda.

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, 5 Agustus 2026



Rendy Ariesta Chandra

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

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

Nama : RENDY ARIESTA CHANDRA
NIM : 41522010139
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : Pemodelan Topik Kontekstual Teks Hadis Kategori Perintah, Larangan, Dan Informasi Menggunakan Bertopic Berbasis Indobert.

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, 05 Agustus 2026

Yang menyatakan,



Rendy Ariesta Chandra

PEMODELAN TOPIK KONTEKSTUAL TEKS HADIS KATEGORI PERINTAH, LARANGAN, DAN INFORMASI MENGUNAKAN BERTOPIC BERBASIS INDOBERT

RENDY ARIESTA CHANDRA

ABSTRAK

Teks hadis merupakan sumber hukum dan panduan ajaran Islam yang kaya pesan kontekstual. Banyaknya teks hadis memerlukan sistem komputasi yang mampu mengelompokkannya secara otomatis untuk mempermudah pemahaman. Penelitian ini bertujuan mengklasifikasikan dan memodelkan topik kontekstual teks hadis ke dalam tiga kategori utama, yaitu perintah, larangan, dan informasi. Definisi variabel penelitian ini meliputi klasifikasi hukum, *coherence score*, dan *topic diversity*. Metode analisis data menerapkan pendekatan hibrida (*hybrid approach*) yang menggabungkan metode berbasis aturan (*rule-based*) Lexicon untuk klasifikasi awal, dilanjutkan metode *State-of-the-Art* BERTopic berbasis IndoBERT untuk tahap pemodelan. Pengambilan sampel menggunakan *total sampling*, dengan dataset berjumlah 12.370 baris teks terjemahan hadis bersumber dari Kitab Shahih Bukhari dan Shahih Muslim. Hasil penelitian membuktikan pelabelan lexicon berhasil memetakan 1.641 data perintah, 1.477 data larangan, dan 9.252 data informasi dengan akurasi 82,67% berdasarkan pengujian 300 sampel validasi manual. Sebelum pemodelan, data melalui pra-pemrosesan teks (*text processing*) yang berfokus pada pembersihan dan penghapusan kata ingkar (*stopword removal*). Pada tahap pemodelan, arsitektur BERTopic IndoBERT secara konsisten menunjukkan performa jauh lebih unggul dibandingkan algoritma LDA (*Latent Dirichlet Allocation*) di ketiga kategori. Pada kategori perintah, BERTopic menghasilkan *coherence score* 0.4811 dan *topic diversity* 0.7912 (LDA: 0.3054 dan 0.2636). Pada kategori larangan, BERTopic memperoleh *coherence score* 0.4677 dan *topic diversity* 0.7947 (LDA: 0.2965 dan 0.3357). Pada kategori informasi, BERTopic meraih *coherence score* 0.4514 dan *topic diversity* 0.7292 (LDA: 0.4610 dan 0.3357). Kesimpulannya, arsitektur BERTopic terbukti jauh lebih koheren dan bervariasi. Penelitian ini diharapkan memberikan kontribusi signifikan dalam pengembangan teknologi keislaman (*Islamic tech*) serta mempermudah pencarian dan pemahaman hadis secara digital.

Kata kunci: Hadis, Lexicon, BERTopic, IndoBERT, LDA

MODELING CONTEXTUAL TOPICS IN HADITH TEXTS IN THE CATEGORIES OF COMMAND, PROHIBITION, AND INFORMATION USING INDOBERT-BASED BERTOPIC

RENDY ARIESTA CHANDRA

ABSTRACT

Hadith texts are a primary Islamic law source rich in contextual messages. The vast hadith volume requires automated computational grouping to facilitate understanding. This study aims to classify and model contextual hadith topics into three main categories: commands, prohibitions, and information. Research variables include legal classification, coherence score, and topic diversity. The analysis applies a hybrid approach combining a rule-based Lexicon method for initial classification and the State-of-the-Art IndoBERT-based BERTopic method for modeling. The sampling technique used is total sampling, utilizing a dataset of 12,370 translated hadith texts from Sahih Bukhari and Sahih Muslim. The results demonstrated that lexicon labeling successfully mapped 1,641 command data, 1,477 prohibition data, and 9,252 information data with 82.67% accuracy based on 300 manually validated samples. Prior to modeling, the data underwent text processing focused on cleaning and stopword removal. In the modeling stage, the BERTopic IndoBERT architecture consistently demonstrated far superior performance compared to the LDA (Latent Dirichlet Allocation) algorithm across all three categories. For commands, BERTopic yielded a 0.4811 coherence score and 0.7912 topic diversity (LDA: 0.3054 and 0.2636). For prohibitions, BERTopic obtained a 0.4677 coherence score and 0.7947 topic diversity (LDA: 0.2965 and 0.3357). For information, BERTopic achieved a 0.4514 coherence score and 0.7292 topic diversity (LDA: 0.4610 and 0.3357). In conclusion, the BERTopic architecture proved to be highly coherent and diverse. This research is expected to contribute significantly to the development of Islamic technology (Islamic tech) and streamline the digital search and understanding of hadiths.

Keywords: *Hadith, Lexicon, BERTopic, IndoBERT, LDA.*

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.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
BAB II TINJAUAN PUSTAKA	6
2.1 Penelitian Terdahulu	6
2.2 Teori Pendukung	12
2.2.1 Konsep Dasar Hadis dan Strukturnya.....	12
2.2.2 Pemrosesan Bahasa Alami (<i>Natural Language Processing</i>) pada Teks Hadis12	

2.2.3	Pra-Pemrosesan Teks (<i>Text Preprocessing</i>).....	13
2.2.4	Klasifikasi Berbasis Aturan (<i>Lexicon-Based</i>)	13
2.2.5	Latent Dirichlet Allocation (LDA).....	13
2.2.6	Arsitektur BERTopic dan Model Bahasa IndoBERT	14
2.2.7	Arsitektur BERTopic dan Model Bahasa IndoBERT	14
BAB III METODE PENELITIAN.....		16
3.1	Jenis Penelitian dan Desain Penelitian.....	16
3.2	Tahapan Penelitian.....	16
3.2.1	Pengumpulan Data Sekunder	18
3.2.2	Konversi Format Data	18
3.2.3	Klasifikasi Awal Berbasis Aturan	18
3.2.4	Percabangan Data (Sub-Dataset).....	18
3.2.5	Pra-Pemrosesan Teks (<i>Text Preprocessing</i>).....	18
3.2.6	Pemodelan Topik Kontekstual (Arsitektur BERTopic)	19
3.2.7	Perancangan Model Baseline Pembanding (LDA).....	20
3.2.8	Penggabungan dan Evaluasi Komparatif.....	20
3.3	Analisis Kebutuhan Lingkungan Eksperimen.....	21
BAB IV HASIL DAN PEMBAHASAN		23
4.1	Analisis Algoritma secara Detil	23
4.1.1	Mekanisme Klasifikasi Kontekstual Berbasis <i>Lexicon Scoring</i>	23
4.1.2	Analisis Pra-Pemrosesan Teks (<i>Text Preprocessing</i>) Terintegrasi.....	25
4.1.3	Arsitektur Pemodelan Topik <i>BERTopic</i> Berbasis <i>IndoBERT</i>	27
4.1.4	Analisis Algoritma <i>Baseline</i> Pembanding (<i>Latent Dirichlet Allocation</i>)	31
4.2	Hasil Pengumpulan dan Pra-Pemrosesan Teks.....	32
4.2.1	Hasil Pengumpulan dan Karakteristik Dataset	32
4.2.2	Implementasi dan Hasil <i>Text Preprocessing</i>	33
4.3	Hasil Klasifikasi Awal Berbasis <i>Lexicon Scoring</i>	35
4.4	Visualisasi dan Pemetaan Tema Topik	37
4.5	Evaluasi Algoritma	52
4.6	Analisis Kesalahan (<i>Error Analysis</i>).....	54
BAB V KESIMPULAN DAN SARAN.....		56
5.1	Kesimpulan	56
5.2	Saran	57
DAFTAR PUSTAKA.....		60
LAMPIRAN.....		64

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	6
Tabel 3.1 Spesifikasi Lingkungan Laboratorium Eksperimen.....	21
Tabel 4.1 Daftar Kata Kunci (Lexicon) Klasifikasi Hukum Kontekstual.....	24
Tabel 4.2 Daftar Parameter Penanda Sabda (Matan Markers)	26
Tabel 4.3 Ilustrasi Transformasi Data dan Angka pada Arsitektur BERTopic	30
Tabel 4.4 Klasifikasi Tambahan Custom Stopwords Teks Hadis.....	34
Tabel 4.5 Contoh Transparansi Perubahan Teks pada Tahap Preprocessing	35
Tabel 4.6 Distribusi Hasil Klasifikasi Kontekstual Matan Hadis	36
Tabel 4.7 Hasil Pemetaan Topik Kontekstual Teks Hadis Perintah menjadi Tema	43
Tabel 4.8 Hasil Pemetaan Topik Kontekstual Teks Hadis Larangan menjadi Tema	45
Tabel 4.9 Hasil Pemetaan Top 20 Topik Kontekstual Teks Hadis Informasi menjadi Tema	50
Tabel 4.10 Hasil Evaluasi Komparatif BERTopic (IndoBERT) dan LDA.....	52



DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Penelitian	17
Gambar 4.1 Visualisasi <i>Topic Word Scores</i> Kategori Perintah	38
Gambar 4.2 Visualisasi <i>Topic Word Scores</i> Kategori Larangan	39
Gambar 4.3 Visualisasi <i>Topic Word Scores</i> Kategori Informasi	40
Gambar 4.4 Intertopic Distance Map - Hadits Informasi	41



DAFTAR LAMPIRAN

Lampiran 1. Kartu Asistensi	64
Lampiran 2. Curriculum Vitae	65
Lampiran 3. Surat Pencatatan Ciptaan	66
Lampiran 4. Sertifikat BNSP	67
Lampiran 5 Tabel Informasi Hasil Pemetaan Komprehensif Topik Hadis.....	68

