



HZKIA: Sebuah model machine learning untuk pemetaan topik penelitian digitalisasi dan jejaring penulis dengan pendekatan sBERT dan HDBSCAN

**TUGAS AKHIR
SKRIPSI**



Hizkia Abqaria

41822010044

**UNIVERSITAS
MERCU BUANA**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**



HZKIA: Sebuah model machine learning untuk pemetaan topik penelitian digitalisasi dan jejaring penulis dengan pendekatan sBERT dan HDBSCAN



**TUGAS AKHIR
SKRIPSI**

**Hizkia Abqaria
41822010044**

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Hizkia Abqaria
NIM : 41822010044
Program Studi : Sistem Informasi

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“HZKIA : sebuah model machine learning untuk pemetaan topik penelitian digitalisasi dan jejaring penulis dengan pendekatan sBERT dan HDBSCAN” 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.

Jakarta, 25 Desember 2025



Hizkia Abqaria.

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 : Hizkia Abqaria
NIM /Student ID Number : 41822010044
Program Studi /Study Program : Sistem Informasi

Dengan Judul Tugas Akhir

/The title:

“HZKIA : sebuah model machine learning untuk pemetaan topik penelitian digitalisasi dan jejaring penulis dengan pendekatan sBERT dan HDBSCAN”

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

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

29 Desember 2025

dengan nilai persentase sebesar :

/with a percentage value of:

10%

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.*

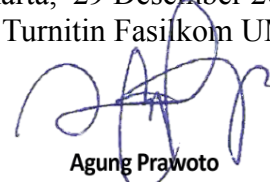
File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1ryHIETunI8MSWmhgPZdw8JtSbxQGKXPn/view?usp=drive_link



Jakarta, 29 Desember 2025
Admin Turnitin Fasilkom UMB



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

HALAMAN PENGESAHAN

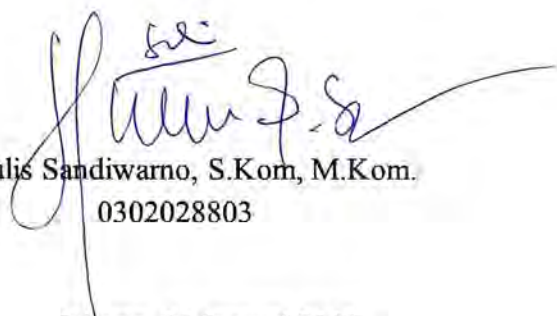
Laporan Skripsi ini diajukan oleh:

Nama : Hizkia Abqaria
NIM : 41822010044
Program Studi : Sistem Informasi
Judul Laporan Skripsi : HZKIA : Sebuah Model Machine Learning untuk Pemetaan Topik Penelitian Digitalisasi dan Jejaring Penulis dengan Pendekatan sBERT dan HDBSCAN

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing,



Sulis Sandiwarno, S.Kom, M.Kom.
0302028803

Jakarta, 25 Januari 2026

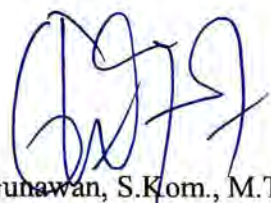
Mengetahui,

Dekan Fakultas Ilmu Komputer

Ketua Program Studi
Sistem Informasi



Dr. Bambang Jokonowo, S.Si., M.T.I.
NIDN: 0320037002



Wawan Gunawan, S.Kom., M.T., M.Kom.
NIDN: 0424108104

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.

Penulis menyadari bahwa tanpa bantuan, dukungan, dan bimbingan dari berbagai pihak, penyusunan tugas akhir ini tidak akan berjalan dengan lancar. Oleh karena itu, penulis menyampaikan terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Prof Dr. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana Jakarta
2. Dr. Bambang Jokonowo, S.Si., M.T.I. selaku Dekan Fakultas Ilmu Komputer
3. Wawan Gunawan, S.Kom, MT., M.Kom. selaku Ketua Program Studi Sistem Informasi
4. Bapak Sulis Sandiwarno, S.Kom., M.Kom. selaku dosen pembimbing tugas akhir yang telah dengan sabar membimbing dan memberikan masukan yang sangat berarti selama proses penyusunan skripsi ini.
5. Mama dr. Mustika dan papa Hafiz, MA. yang sudah mendukung penuh segala doa, dukungan moral dan materiil, serta kasih sayang yang tidak pernah putus.
6. Kakak-ku Haiza Mutia Hani, S.Psi , adik-ku Hani Dza Izzatul Mecca dan adik-ku Muhammad Azizul Manal, yang dengan penuh pengertian telah memberikan dukungan selama proses penyusunan skripsi, tidak mengganggu penulis saat fokus mengerjakan tugas akhir, serta dengan sabar menghadapi penulis meskipun sering menjadi sasaran keisengan ketika penulis sedang merasa penat.
7. Para dosen dan staf TU khususnya pak Ajat, pak Erfan, dan pak Agung yang telah membantu dalam proses pemberkasan dan pendaftaran sidang ini.
8. Nanda, Fira, Nadhira, Thania, Bulan, Sheela, dan Setia, yang telah memberikan semangat, motivasi, serta kebersamaan yang menyenangkan selama masa kuliah maupun pengerjaan tugas akhir, dan selalu membuat penulis tertawa.
9. Tim People analytic and enabler management PT Telekomunikasi Selular, khususnya mbak Dyas dan mas Sule selaku mentor, yang dengan penuh pengertian telah memberikan kemudahan izin untuk bimbingan, sehingga proses persiapan proposal tugas akhir hingga penyusunan tugas akhir dapat berjalan dengan baik, serta atas segala ilmu, inspirasi dan bimbingan yang telah diberikan kepada penulis.
10. Tim Microsoft Teneleven PT Media Kreasi Asia, khususnya tim Data Engineer, yang telah membimbing penulis dalam memahami lingkungan dan alur kerja profesional di dunia industri, sehingga penulis mampu menyeimbangkan kegiatan pekerjaan dengan

penyusunan tugas akhir tanpa kendala berarti. Terima kasih atas ilmu, pengalaman, serta bimbingan yang sangat berharga bagi penulis.

11. Teman-teman Sistem Informasi 2022 Menteng–Meruya–Warbun, Sistem Informasi 2021 dan 2023 Meruya, serta HIMSISFO UMB periode 2024/2025 dan periode 2023/2024, atas pengalaman, kebersamaan, dan pembelajaran yang sangat berharga, yang telah membentuk penulis hingga menjadi seperti saat ini.
12. Teknologi kecerdasan buatan, khususnya ChatGPT dan Claude, yang telah membantu penulis dalam mempermudah proses pemahaman, eksplorasi ide, serta penyusunan materi selama pengerjaan tugas akhir.

Jakarta, 25 Desember 2025

Penulis



HALAMAN PENYATAAN PERSTUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

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

Nama Mahasiswa : Hizkia Abqaria
NIM : 41822010044
Fakultas/ Program studi : Ilmu Komputer / Sistem Informasi
Judul Tugas Akhir : HZKIA: Sebuah model machine learning untuk pemetaan topik penelitian digitalisasi dan jejaring penulis dengan pendekatan sBERT dan HDBSCAN

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.

Jakarta, 25 Desember 2025

Yang menyatakan,



Hizkia Abqaria

HZKIA: Sebuah model machine learning untuk pemetaan topik penelitian digitalisasi dan jejaring penulis dengan pendekatan sBERT dan HDBSCAN

ABSTRAK

Perkembangan penelitian dalam ruang lingkup digitalisasi yang pesat memerlukan metode pemetaan topik yang lebih akurat dan komprehensif untuk memahami tren dan dinamika kolaborasi peneliti. Beberapa penelitian sebelumnya sudah melakukan evaluasi terhadap pemetaan topik dengan menggunakan beberapa pendekatan, seperti LDA, K-Means, dan BERTopic. Akan tetapi, metode tersebut memiliki beberapa keterbatasan dalam menghasilkan topik yang stabil, memerlukan pelabelan manual, dan mengabaikan dimensi kolaborasi antar peneliti. Oleh karena itu dalam penelitian ini kami mengusulkan sebuah model yang bernama Hybridized Zoning of Knowledge via Integrated Analysis (HZKIA), sebuah pendekatan baru yang mengintegrasikan sBERT untuk pembuatan embedding, UMAP untuk reduksi dimensi, dan HDBSCAN yang dioptimasi menggunakan optuna untuk clustering. Penamaan topik dilakukan secara otomatis menggunakan KeyBERT dengan teknik MMR. Model ini juga mengintegrasikan analisis jejaring sosial untuk melihat hubungan antar peneliti dan memvisualisasikannya dalam bentuk knowledge graph. Hasil penelitian menunjukkan bahwa model HZKIA mampu menghasilkan pemetaan topik yang lebih akurat, stabil, dan kontekstual dibandingkan metode sebelumnya, diukur dari nilai Coherence Score, Topic Diversity, dan Silhouette Score yang lebih stabil. Model ini juga berhasil menampilkan hubungan sosial dan kolaborasi dalam komunitas akademik secara komprehensif, memberikan pemahaman yang lebih mendalam tentang ekosistem penelitian digitalisasi.

MERCU BUANA

Kata kunci : Topic modelling, HZKIA, sBERT, digitalisasi, Social network analysis, knowledge graph, text mining

HZKIA: A Machine Learning Model for Mapping Digitalization Research Topics and Author Network Using an sBERT and HDBSCAN Approach

ABSTRACT

The rapid growth of research in the field of digitalization necessitates more accurate and comprehensive topic mapping methods to understand research trends and the dynamics of scholarly collaboration. Several previous studies have evaluated topic modeling approaches using methods such as LDA, K-Means, and BERTopic. However, these methods exhibit several limitations, including unstable topic generation, reliance on manual labeling, and the neglect of collaboration dimensions among researchers. To address these limitations, this study proposes a novel model named Hybridized Zoning of Knowledge via Integrated Analysis (HZKIA), which integrates sBERT for embedding generation, UMAP for dimensionality reduction, and HDBSCAN optimized using Optuna for clustering. Topic labeling is performed automatically using KeyBERT with the Maximal Marginal Relevance (MMR) technique. Furthermore, the proposed model incorporates social network analysis to examine relationships among researchers and visualizes them in the form of a knowledge graph. The experimental results demonstrate that the HZKIA model produces more accurate, stable, and conceptually coherent topic mappings compared to existing methods, as evidenced by more stable values of Coherence Score, Topic Diversity, and Silhouette Score. In addition, the model successfully reveals social relationships and collaboration patterns within the academic community in a comprehensive manner, providing deeper insights into the digitalization research ecosystem.

Keywords: *Topic modeling, HZKIA, sBERT, digitalization, social network analysis, knowledge graph, text mining*

DAFTAR ISI

HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN UJI TURNITIN	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PENYATAAN PERSTUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB	vii
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	3
1.5 Batasan Penelitian.....	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	7
2.1 Digitalisasi	7
2.2. Jejaring dan Relasi Kolaboratif penulis	8
2.3 <i>Machine Learning</i> untuk <i>topic modelling</i>	9
2.3.1 Embedding	11
2.3.2 Dimensionality reduction.....	12
2.3.3 Clustering.....	13
2.3.4 <i>Hyperparameter Optimization</i>	14
2.4 Penamaan <i>cluster</i> otomatis	14
2.5 <i>Evaluation Matrix</i>	15
2.6 Penelitian Terdahulu	16
2.7 Analisis Literature Review	36
BAB III METODOLOGI PENELITIAN	38
3.1 Jenis Penelitian.....	38
3.2 Tahapan Penelitian.....	38
3.2.1 <i>Problem definition</i>	39
3.2.2. <i>Data Collection</i>	39

3.2.3 Data Processing.....	40
3.2.4 HZKIA modelling	42
BAB IV HASIL DAN PEMBAHASAN	59
4.1 <i>Experimental Setup</i>	59
4.2 <i>Dataset</i>	60
4.3 <i>Parameter Setting</i>	62
4.4 Hasil	65
4.4.1 Struktur Topik yang terbentuk	65
4.4.2 Visualisasi Struktur <i>Spatial</i> Topik	67
4.4.3 <i>Evaluation</i> model HZKIA.....	68
4.4.4. Analisis Dokumen <i>Noise</i>	70
4.4.5 Hubungan Antar Topik	72
4.5 Ablation Study	74
4.6 Diskusi	78
4.7 Implikasi Penelitian	82
4.7.1 Implikasi Metodologis	82
4.7.2 Implikasi Praktis	83
4.7.3 Implikasi untuk pengembangan riset masa Depan.....	84
BAB V KESIMPULAN DAN SARAN	86
5.1 Kesimpulan	86
5.2 Saran	87
DAFTAR PUSTAKA	88
LAMPIRAN.....	92

DAFTAR TABEL

Tabel 2.1 Literature review	16
Tabel 3.1 Library extract data from PDF to CSV	41
Tabel 3.2 distribusi hasil clustering HDBSCAN	51
Tabel 3.3 distribusi hasil pelabelan otomatis KeyBERT	56
Tabel 4.1 distribusi tahun dataset.....	60
Tabel 4.2 sample trial optuna.....	63
Tabel 4.3 final parameter model HZKIA.....	64
Tabel 4.4 distribusi dokumen dan keyword per topik.....	65
Tabel 4.5 evaluation matrix cluster quality HZKIA model	68
Tabel 4.6 analisis dokumen noise	70
Tabel 4.7 topic modelling comparison.....	78



DAFTAR GAMBAR

Gambar 2. 1 Topology machine learning	9
Gambar 2. 2 hubungan AI ML NLP DL	10
Gambar 2. 3 perbedaan arsitektur BERT dan sBERT	11
Gambar 3.1 flow research methodology	38
Gambar 3.2 Data Processing flow	40
Gambar 3.3 Arsitektur sEBRT	43
Gambar 3.4 HDBSCAN flow	50
Gambar 4.1 visualisasi density map.....	67
Gambar 4.2 Knowledge Graph Cluster.....	73
Gambar 4.3 Compherensive waterfall analysis	75
Gambar 4.4 component impact heatmap per matric	77
Gambar 4.5 Topic model comparison.....	78
Gambar 4.6 Author relation	80

