



**PEMETAAN TREN PENELITIAN AKADEMIK
BERDASARKAN JUDUL PUBLIKASI MENGGUNAKAN
WORD2VEC DAN K-MEANS CLUSTERING**

**TUGAS AKHIR
SKRIPSI**

BIMO BAGAS RIYANTO

41522010049

**UNIVERSITAS
MERCU BUANA**

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



**PEMETAAN TREN PENELITIAN AKADEMIK
BERDASARKAN JUDUL PUBLIKASI MENGGUNAKAN
WORD2VEC DAN K-MEANS CLUSTERING**

**TUGAS AKHIR
SKRIPSI**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

**UNIVERSITAS
MERCU BUANA**
BIMO BAGAS RIYANTO
41522010049

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

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Bimo Bagas Riyanto

NIM : 41522010049

Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“Pemetaan Tren Penelitian Akademik Berdasarkan Judul Publikasi Menggunakan Word2Vec dan K-Means Clustering” 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, 30 Juli 2026



Bimo Bagas Riyanto

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 : Bimo Bagas Riyanto
NIM /Student ID Number : 41522010049
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

The title:

“Pemetaan Tren Penelitian Akademik Berdasarkan Judul Publikasi Menggunakan Word2Vec dan K-Means Clustering”

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

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

17 Juli 2026

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.*

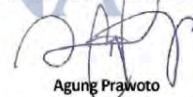
File hasil cek similarity turnitin:

Turnitin similarity report file

https://drive.google.com/file/d/1S4fymGm9v3TNe_OkayVrks0iQIC_xAW/view?usp=drive_link



Jakarta, 17 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

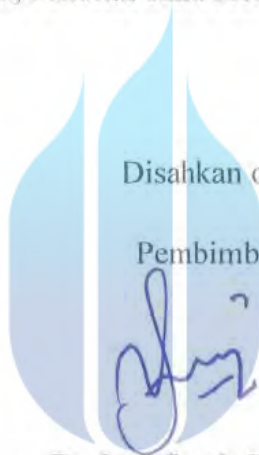
Tugas Akhir ini diajukan oleh:

Nama : BIMO BAGAS RIYANTO
NIM : 41522010049
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : Pemetaan Tren Penelitian Akademik Berdasarkan Judul Publikasi Menggunakan Word2Vec dan K-Means Clustering

Telah berhasil dipertahankan pada sidang tanggal 30 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

MERCU BUANA

Jakarta, 30 Juli 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 kehadirat Tuhan yang Maha Esa, atas segala rahmat dan ridha-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang merupakan salah satu persyaratan kelulusan Program Studi Strata Satu (S1) pada jurusan Teknik Informatika pada Fakultas Ilmu Komputer, Universitas Mercu Buana.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna, karena kesempurnaan sejatinya hanya milik Tuhan yang Maha Esa. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Serta berkat dukungan, motivasi, bantuan, bimbingan, dan doa dari banyak pihak, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., M.T.I selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
4. 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 telah menjadi sumber kekuatan, doa, dukungan tak terhingga, serta segala pengorbanan, kasih sayang, dan kepercayaan yang telah diberikan, sehingga penulis mampu melewati setiap tantangan dan menyelesaikan pendidikan hingga tahap ini dengan baik.
6. Anggrina Aufa A. selaku partner penulis yang telah memberikan dukungan, perhatian, motivasi, dan semangat yang selalu diberikan selama proses penyusunan Tugas Akhir ini.
7. Teman-teman seperjuangan, yang telah menjadi tempat berbagi cerita, semangat, dan dukungan selama 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 Laporan Tugas Akhir ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 30 Juli 2026

Bimo Bagus Riyanto



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

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

Nama : BIMO BAGAS RIYANTO
NIM : 41522010049
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : Pemetaan Tren Penelitian Akademik Berdasarkan Judul Publikasi Menggunakan Word2Vec dan K-Means Clustering

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.

MERCU BUANA

Jakarta, 30 Juli 2026

Yang menyatakan,



Bimo Bagus Riyanto

PEMETAAN TREN PENELITIAN AKADEMIK BERDASARKAN JUDUL PUBLIKASI MENGGUNAKAN WORD2VEC DAN K-MEANS CLUSTERING

BIMO BAGAS RIYANTO

ABSTRAK

Institusi akademik menghadapi tantangan dalam memetakan distribusi dan tren topik penelitian dosen secara sistematis. Penelitian ini mengusulkan pendekatan yang menggabungkan word embedding Word2Vec (arsitektur CBOW, dimensi 100) dengan K-Means Clustering untuk pemodelan topik otomatis terhadap 2.107 judul publikasi dosen Fakultas Ilmu Komputer, Universitas Y, dikumpulkan melalui ekstraksi otomatis dari Google Scholar. Pendekatan ini dipilih karena mampu menangkap hubungan semantik antar istilah akademik, berbeda dengan TF-IDF yang hanya mengandalkan frekuensi leksikal. Jumlah kluster optimal ($k=4$) ditentukan menggunakan Metode Elbow dan divalidasi melalui Silhouette Score, menghasilkan nilai 0,3022 dari 1.946 judul unik yang terbagi ke dalam empat cluster (1.068, 637, 223, dan 18 dokumen) dengan topik dominan: pengembangan aplikasi dan model berbasis data, perancangan sistem informasi dan sistem pendukung keputusan, penerapan machine learning untuk klasifikasi dan prediksi, serta cybersecurity dan sistem paralel. Program Studi Sistem Informasi lebih dominan pada topik perancangan sistem, sementara Teknik Informatika lebih dominan pada machine learning. Analisis tren tahun 1983–2026 menunjukkan titik balik pertumbuhan pesat pada 2017–2019, dengan cluster aplikasi dan sistem informasi konsisten mendominasi, cluster machine learning tumbuh signifikan sejak 2018, dan cluster cybersecurity tetap jarang muncul sepanjang periode pengamatan.

Kata Kunci: pemetaan tren penelitian; word2vec; k-means clustering; natural language processing; analisis judul publikasi

**PEMETAAN TREN PENELITIAN AKADEMIK BERDASARKAN JUDUL
PUBLIKASI MENGGUNAKAN WORD2VEC DAN K-MEANS
CLUSTERING**

BIMO BAGAS RIYANTO

ABSTRACT

Academic institutions face increasing challenges in systematically mapping the distribution and trends of faculty research topics. This study proposes an approach that combines Word2Vec word embedding (CBOW architecture, 100 dimensions) with K-Means Clustering for automatic topic modeling of 2,107 publication titles from lecturers at the Faculty of Computer Science, University Y, collected through automated extraction from Google Scholar. This approach was chosen for its ability to capture semantic relationships between academic terms, in contrast to TF-IDF, which relies solely on lexical frequency. The optimal number of clusters ($k=4$) was determined using the Elbow Method and validated through the Silhouette Score, yielding a value of 0.3022 from 1,946 unique titles distributed across four clusters (1,068, 637, 223, and 18 documents), with dominant topics including application and data-based model development, information system and decision support system design, machine learning applications for classification and prediction, and cybersecurity and parallel systems. The Information Systems study program showed greater dominance in system design topics, while Informatics Engineering was more dominant in machine learning applications. Trend analysis from 1983 to 2026 revealed a sharp turning point in growth between 2017 and 2019, with the application development and information systems clusters consistently dominating, the machine learning cluster showing significant growth since 2018, and the cybersecurity cluster remaining sparsely represented throughout the observation period.

Keywords: research trend mapping; word2vec; k-means clustering; natural language processing; publication title analysis

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	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah Penelitian	4
BAB II TINJAUAN PUSTAKA	6
2.1 Teori Utama	6
2.1.1 Clustering Tren Penelitian Akademik	6
2.1.2 Word Embedding Word2Vec	6
2.1.3 Algoritma K-Means.....	7
2.2 Teori Pendukung.....	8
2.2.1 Python.....	9
2.2.2 Data Mining.....	9
2.2.3 Silhouette Score.....	9
2.3 Penelitian Terdahulu	10
2.4 Gap Penelitian	19
BAB III METODE PENELITIAN	21

3.1	Pendekatan Penelitian	21
3.2	Tahapan Penelitian.....	22
BAB IV PEMBAHASAN		26
4.1	Dataset Penelitian.....	26
4.2	Document Translation.....	27
4.3	Data Understanding	28
4.3.1	Exploratory Data Analysis (EDA)	28
4.3.2	Analisis Data Duplikat dan Data Hilang	31
4.4	Data Preprocessing.....	32
4.4.1	Penanganan Nilai Hilang dan Data Duplikat.....	32
4.4.2	Data Transformation.....	33
4.4.3	Data Cleaning	34
4.4.4	Tokenization	35
4.4.5	Stop Word Removal dan Stemming.....	35
4.4.6	Word Embedding Word2Vec.....	37
4.5	Modeling.....	39
4.5.1	Menghitung Inersia.....	40
4.5.2	Elbow Method	42
4.5.3	K-Means Clustering	43
4.6	Evaluation	45
4.7	Analisis Hasil Clustering	46
4.7.1.	Visualisasi Cluster.....	46
4.7.2.	Analisis Topik Penelitian Dosen	49
4.7.3.	Distribusi Cluster per Program Studi	51
4.7.4.	Analisis Tren Penelitian Berdasarkan Hasil Cluster	53
BAB V KESIMPULAN DAN SARAN.....		55
5.1	Kesimpulan	55
5.2	Saran	56
DAFTAR PUSTAKA.....		57
LAMPIRAN.....		60

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	10
Tabel 4.1 Sampel Dataset Judul Publikasi	26
Tabel 4.2 Sampel Judul Publikasi Sebelum dan Sesudah Translasi	28
Tabel 4.3 Hasil Pengecekan Data	29
Tabel 4.4 Hasil Penanganan Nilai Hilang.....	33
Tabel 4.5 Hasil Proses Tokenisasi	35
Tabel 4.6 Hasil Penghapusan Stop Word.....	37
Tabel 4.7 Perbandingan Silhouette Score di Titik Elbow	43
Tabel 4.8 Hasil Distribusi Jumlah Dokumen per Cluster	44
Tabel 4.9 Hasil Silhouette Score.....	45



DAFTAR GAMBAR

Gambar 2.1 Algoritma K-means	8
Gambar 3.1 Tahapan Penelitian	22
Gambar 4.1 Code Pengecekan Struktur Data.....	29
Gambar 4.2 Visualiasi Distribusi Persebaran Publikasi per Tahun	30
Gambar 4.3 Visualisasi Perbandingan Publikasi	30
Gambar 4.4 Code Pengecekan Data Duplikat dan Nilai Hilang	31
Gambar 4.5 Hasil Pengecekan Data Duplikat dan Nilai Hilang.....	32
Gambar 4.6 Code Penanganan Nilai Hilang	33
Gambar 4.7 Code Data Transformasi	34
Gambar 4.8 Code Data Cleaning	34
Gambar 4.9 Code Tokenisasi	35
Gambar 4.10 Code Stop Word Removal dan Stemming	36
Gambar 4.11 Code Konfigurasi Parameter dan Pelatihan Word2Vec.....	38
Gambar 4.12 Code Agregasi Vektor Kata Menjadi Dokumen Vektor	39
Gambar 4.13 Hasil Pembuatan Matriks Vektor Dokumen	39
Gambar 4.14 Code Perhitungan Inersia	40
Gambar 4.15 Hasil Perhitungan Inersia	41
Gambar 4.16 Visualisasi Elbow Method	42
Gambar 4.17 Code Implementasi K-Means.....	44
Gambar 4.18 Code Implementasi Silhouette Score	45
Gambar 4.19 Visualisasi Cluster Topik Penelitian Dosen dengan PCA	47
Gambar 4.20 Visualisasi Word Cloud per Cluster.....	50
Gambar 4.21 Distribusi Cluster per Program Studi	52
Gambar 4.22 Tren Penelitian per Cluster dari Tahun ke Tahun	53

DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi	60
Lampiran 2 Curriculum Vitae	61
Lampiran 3 Surat Pernyataan & Pengalihan HAKI	63
Lampiran 4 Bukti BNSP	65

