

Kombinasi *BERT* dan *Random Forest* untuk Sistem Rekomendasi

Dosen Pembimbing Berdasarkan Analisis Topik dan

kata kunci Judul Skripsi Mahasiswa

SKRIPSI

DISUSUN OLEH

ZUKH ATIKA HARAHAH

NPM. 2209020214



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

Kombinasi *BERT* dan *Random Forest* untuk Sistem Rekomendasi

Dosen Pembimbing Berdasarkan Analisis Topik dan

kata kunci Judul Skripsi Mahasiswa

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Teknologi Informasi, pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

ZUKH ATIKA HARAHAHAP

NPM. 2209020214

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN**

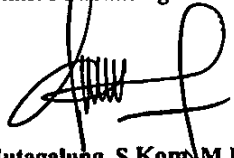
2026

LEMBAR PENGESAHAN

Judul Skripsi : Kombinasi *BERT* dan *Random Forest* untuk Sistem
Rekomendasi Dosen Pembimbing Berdasarkan Analisis
Topik dan kata kunci Judul Skripsi Mahasiswa

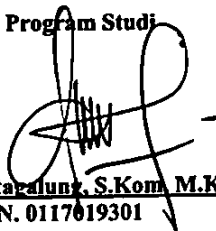
Nama Mahasiswa : Zukh Atika Harahap
NPM : 2209020214
Program Studi : Teknologi Informasi

Menyetujui
Komisi Pembimbing



(Fatma Sari Hutagalung, S.Kom, M.Kom)
NIDN. 0117019301

Ketua Program Studi



(Fatma Sari Hutagalung, S.Kom, M.Kom)
NIDN. 0117019301

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

Kombinasi *BERT* dan *Random Forest* untuk Sistem Rekomendasi

Dosen Pembimbing Berdasarkan Analisis Topik dan

kata kunci Judul Skripsi Mahasiswa

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 20 Januari 2026

Yang membuat pernyataan



Zukh Atika Harahap

Npm. 2209020214

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : ZUKH ATIKA HARAHAAP
NPM : 2209020214
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul: Kombinasi *BERT* dan *Random Forest* untuk Sistem Rekomendasi Dosen Pembimbing Berdasarkan Analisis Topik dan kata kunci Judul Skripsi Mahasiswa

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 20 Januari 2026

Yang membuat pernyataan



Zukh Atika Harahap

NPM. 2209020214

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Zukh Atika Harahap
Tempat dan Tanggal Lahir : Sialambue, 03 Juni 2003
Alamat Rumah : Desa Sialambue
Telepon/Faks/HP : 081262605657
E-mail : zukhatikaharahap2@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SDN 0105 SIBUHUAN TAMAT: 2016
SMP : MTsN 1 Padang Lawas TAMAT: 2019
SMA : SMA 1 Barumun TAMAT: 2022

KATA PENGANTAR



Puji syukur ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Kombinasi BERT dan Random Forest untuk Sistem Rekomendasi Dosen Pembimbing Berdasarkan Analisis Topik dan Kata Kunci Judul Skripsi Mahasiswa” dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara. Penulis menyadari bahwa dalam proses penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada yang terhormat :

1. Bapak Prof. Dr. Akrim, M.Pd., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU).
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firaahmi Rizky, S.Kom., M.Kom selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi.
4. Bapak Mhd. Basri, S.Si., M.Kom selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi.
5. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom selaku Ketua Program Studi Teknologi Informasi sekaligus Pembimbing yang dengan penuh kesabaran memberikan pengarahan, bimbingan, dan motivasi kepada penulis.
6. Bapak Okvi Nugroho, S.Kom., M.Kom selaku sekretaris Program Studi Teknologi Informasi yang selalu memberikan arahan dan bimbingan.

7. Ayahanda penulis, Yung Yung, S.Pd, yang merupakan cinta pertama penulis serta selalu memberikan doa, dukungan, dan semangat.
8. Ibunda penulis, Khairani, S.Pd, yang merupakan pintu surga bagi penulis serta dengan kasih sayang, doa, dan pengorbanan selalu menjadi sumber kekuatan dalam setiap langkah.
9. Saudara/i penulis, yaitu Putri Sakinah Harahap, S.Pd, Muhammad Zulhammuddin Harahap, S.Kom, Rizki Faurina Harahap, S.Agr, dan Hilliyatul Aulia Harahap, yang senantiasa memberikan dukungan moral, semangat, dan doa.
10. Terima kasih juga penulis sampaikan kepada NPM 2209020198 yang telah membantu dan menemani penulis selama proses penyusunan skripsi ini, serta memberikan dukungan dalam setiap tahapnya.
11. Penulis juga mengucapkan terima kasih kepada teman-teman yang selalu hadir memberikan tawa, semangat, bantuan, dan kebersamaan yang tidak tergantikan selama masa perkuliahan hingga pengerjaan skripsi ini.
12. Terima kasih penulis sampaikan kepada diri sendiri, yang telah bertahan, berjuang, dan tidak menyerah meskipun tidak mudah, serta terus melangkah sampai pada titik ini.
13. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

Kombinasi BERT dan Random Forest untuk Sistem Rekomendasi

Dosen Pembimbing Berdasarkan Analisis Topik dan

Kata Kunci Judul Skripsi Mahasiswa

ABSTRAK

Pemilihan dosen pembimbing skripsi merupakan proses penting dalam penyelesaian studi mahasiswa, namun di banyak perguruan tinggi penentuan pembimbing masih dilakukan secara manual sehingga berpotensi menimbulkan ketidaktepatan antara topik penelitian mahasiswa dan bidang keahlian dosen. Penelitian ini bertujuan mengembangkan sistem rekomendasi dosen pembimbing berbasis analisis semantik judul skripsi menggunakan Bidirectional Encoder Representations from Transformers (BERT) serta algoritma Random Forest. BERT digunakan untuk mengekstraksi fitur semantik dari judul skripsi sehingga menghasilkan representasi vektor berdimensi tinggi yang mampu menangkap konteks dan makna topik secara lebih tepat. Representasi vektor tersebut kemudian diproses menggunakan Random Forest untuk menentukan dosen pembimbing yang relevan sesuai bidang keilmuan dosen yang tersedia pada data latih. Sistem yang dibangun diharapkan dapat membantu proses penentuan pembimbing secara lebih objektif, terstruktur, dan efisien, serta mengurangi ketergantungan pada penunjukan manual yang bersifat subjektif.

Kata Kunci: BERT; Random Forest; Sistem Rekomendasi; Pembimbing Skripsi.

*Combination of BERT and Random Forest for Thesis Supervisor
Recommendation System Based on Topic Analysis and
Keyword Extraction from Students' Thesis Titles*

ABSTRACT

The selection of thesis supervisors is an essential process in academic completion; however, in many universities, this process is still performed manually, potentially causing misalignment between students' research topics and lecturers' academic fields. This research aims to develop a supervisor recommendation system based on the semantic analysis of thesis titles using Bidirectional Encoder Representations from Transformers (BERT) and the Random Forest algorithm. BERT is utilized to extract semantic features from thesis titles, producing high-dimensional vector representations that capture contextual meaning more accurately. These vector representations are then processed by Random Forest to determine relevant supervisors according to the academic expertise contained in the training data. The proposed system is expected to support a more objective, structured, and efficient supervisor assignment process, reducing dependency on manual and subjective decision-making.

Keywords: *BERT; Random Forest; Recommendation System; Thesis Supervisor.*

DAFTAR ISI

LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS.....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI.....	iv
RIWAYAT HIDUP	v
KATA PENGANTAR.....	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
BAB I. PENDAHULUAN.....	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	6
1.5. Manfaat Penelitian.....	6
BAB II. LANDASAN TEORI	7
2.1. Machine Learning.....	7
2.2. Bidirectional Encoder Representations From Transformer (BERT)....	8
2.3. Random Forest	12
2.4. Sistem Rekomendasi Dosen Pembimbing.....	13
2.5. Penelitiin Terdahulu	15
2.6. Analisis GAP.....	19
BAB III. METODOLOGI PENELITIAN	21
3.1. Alur Penelitian.....	21
3.2. Diagram Rancangan Sistem dan Pemodelan UML	22
3.3. Sumber dan Jenis data	24
3.4. Variabel Penelitian	25
3.5. Tahap Ekstraksi Fitur Menggunakan BERT	26
3.6. Tahap Klasifikasi Menggunakan Random Forest	28
3.7. Rancangan Desain Interface Sistem	30
3.8. Jadwal Penelitian	34
BAB IV. HASIL DAN PEMBAHASAN	35
4.1. Hasil.....	35
4.2. Hasil Ekstraksi Fitur Menggunakan BERT	42
4.3. Hasil Klasifikasi Berdasarkan Embedding BERT Menggunakan Random Forest	42
4.4. Evaluasi Kinerja Model.....	51
4.5. Pembahasan	53
BAB V. KESIMPULAN DAN SARAN	55
5.1. KESIMPULAN	55
5.2. SARAN	56
DAFTAR PUSTAKA	57

DAFTAR TABEL

	HALAMAN	
Tabel 2.1.	Ringkasan Penelitian Terdahulu	15
Tabel 3.1	Representasi Fitur (Embedding BERT)	28
Tabel 3.2	Jadwal Penelitian	34
Tabel 4.1	Hasil Ekstraksi Fitur Menggunakan BERT	40
Tabel 4.2	Tabel Fitur	43
Tabel 4.3	Tabel Input ke Random Forest	43
Tabel 4.4	Tabel Prediksi	43
Tabel 4.5	Hasil Klasifikasi Random Forest Berdasarkan Embedding BERT	44
Tabel 4.6	Hasil Confusion Matrix	52

DAFTAR GAMBAR

		HALAMAN
Gambar 2.1	Ukuran BERT _{BASE} dan BERT _{LARGE}	9
Gambar 2.2	Representasi Teks Sekuensual pada BERT	11
Gambar 2.3	Ilustrasi <i>Random Forest</i>	12
Gambar 3.1	Alur Penelitian	21
Gambar 3.2	Diagram Rancangan Sistem	22
Gambar 3.3	<i>Use Case Diagram</i>	23
Gambar 3.4	<i>Activity Diagram</i>	24
Gambar 3.5	Flowchart BERT	27
Gambar 3.6	Flowchart Random Forest	29
Gambar 3.7	Desain Interface Menu Login Admin	30
Gambar 3.8	Desain Interface Menu utama admin	31
Gambar 3.9	Desain Interface Riwayat Rekomendasi	31
Gambar 3.10	Desain Interface Menu Logout	32
Gambar 3.11	Desain Halaman Input Judul Skripsi Mahasiswa	32
Gambar 3.12	Desain Hasil Rekomendasi Dosen Pembimbing	33
Gambar 4.1	Menu Login	36
Gambar 4.2	Menu Utama	36
Gambar 4.3	Halaman Kelola Dosen	37
Gambar 4.4	Halaman Riwayat Rekomendasi	37
Gambar 4.5	Halaman Logout	38
Gambar 4.6	Halaman Input Judul	38
Gambar 4.7	Halaman Hasil Input Judul	39

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kompetisi global yang semakin ketat di era modern memaksa pendidikan tinggi untuk terus meningkatkan performa dan daya saingnya. Tantangan tersebut tidak hanya terkait dengan kualitas pengajaran dan penelitian, tetapi juga pada efisiensi proses akademik yang melibatkan mahasiswa dan dosen. Untuk mendapatkan gelar sarjana strata satu, mahasiswa program sarjana harus menyelesaikan tugas akhir (Andriani & Wibawanta, 2020).

Di lingkungan perguruan tinggi, baik di tingkat lokal maupun nasional, proses penentuan dosen pembimbing skripsi adalah langkah yang sangat penting. Keberhasilan dan ketepatan waktu studi mahasiswa sering kali bergantung pada kesesuaian antara topik penelitian yang diangkat dengan bidang keahlian dosen pembimbing. Mahasiswa seringkali diwajibkan menyelesaikan tugas akhir di bawah bimbingan satu atau lebih dosen. Penunjukan dosen pembimbing bukanlah tugas yang mudah karena topik tugas akhir harus sesuai dengan bidang keahlian calon pembimbing (Wijanto et al., 2020).

Fenomena nyata menunjukkan bahwa banyak institusi masih mengandalkan proses manual atau penunjukan yang bersifat subjektif. Misalnya, penugasan dapat diberikan hanya berdasarkan ketersediaan waktu dosen atau berdasarkan kedekatan personal mahasiswa. Akibatnya, sering terjadi ketidakseimbangan beban bimbingan beberapa dosen terlalu banyak membimbing sementara yang lain sedikit dan yang lebih krusial, terjadi ketidakcocokan topik yang mendalam.

Ketidakcocokan ini dapat memperlambat proses bimbingan dan bahkan menurunkan mutu penelitian akhir mahasiswa.

Saat ini, perguruan tinggi memiliki banyak data judul skripsi dan publikasi dosen. Data ini seharusnya dapat digunakan untuk pencocokan yang lebih baik. Namun, upaya-upaya pencocokan otomatis yang sudah ada sering kali menggunakan metode sederhana, seperti mencocokkan kata kunci yang sama (metode TF-IDF). Masalahnya, kata-kata yang berbeda bisa memiliki makna yang sama (sinonim), dan satu kata bisa memiliki banyak arti (polisemi). Metode lama ini gagal memahami makna sesungguhnya (konteks semantik) dari judul skripsi yang singkat. Misalnya, judul tentang "Kecerdasan Buatan" mungkin tidak mencantumkan kata "Machine Learning", padahal keduanya memiliki topik yang sangat relevan. Kesenjangan penelitian (Research Gap) muncul karena kita membutuhkan metode yang tidak hanya mencocokkan kata, tetapi juga mencocokkan makna secara akurat.

Permasalahan dalam penyelesaian tugas akhir, yang sering kali berkontribusi terhadap keterlambatan kelulusan mahasiswa, kerap diakibatkan oleh ketidaksesuaian antara topik penelitian mahasiswa dan keahlian dosen pembimbing. Oleh karena itu diperlukan metode untuk mengatasi ketidaksesuaian tersebut dalam menyelesaikan tugas akhir (No et al., 2024).

Untuk mengatasi masalah pemahaman makna ini, penelitian ini melihat pada perkembangan terbaru di bidang pemrosesan bahasa alami (NLP), yaitu model BERT (Bidirectional Encoder Representations from Transformers). BERT memiliki keunggulan luar biasa karena ia bisa membaca dan memahami teks secara keseluruhan, dari depan dan belakang, untuk menangkap konteks mendalam dari

judul skripsi. Dengan kata lain, BERT membantu kita mengubah judul skripsi menjadi kode makna (vektor) yang sangat kaya.

Setelah kode makna ini didapatkan, kita membutuhkan algoritma yang kuat untuk menggunakannya dalam proses klasifikasi. Dipilihlah algoritma *Random Forest* (RF) karena algoritma ini terkenal kokoh, akurat, dan sangat baik dalam membuat keputusan prediksi berdasarkan banyak fitur yang kompleks. Urgensi penelitian ini adalah membuktikan bahwa kombinasi BERT (untuk memahami makna) dan *Random Forest* (untuk prediksi akurat) dapat menciptakan sistem rekomendasi yang sangat objektif dan efisien.

Untuk mengatasi kelemahan metode pencocokan judul skripsi yang hanya berfokus pada kesamaan kata, penelitian ini menggunakan model *Bidirectional Encoder Representations from Transformers* (BERT). Model ini mampu memahami konteks judul secara menyeluruh karena memproses teks dari dua arah, sehingga judul skripsi mahasiswa dapat direpresentasikan dalam bentuk vektor yang mencerminkan makna topik penelitian secara lebih tepat, meskipun menggunakan istilah yang berbeda.

Hasil representasi dari BERT kemudian diproses menggunakan algoritma *Random Forest* sebagai metode klasifikasi. *Random Forest* dipilih karena mampu menghasilkan keputusan yang lebih stabil dan akurat melalui penggabungan beberapa pohon keputusan, serta efektif dalam mengolah fitur berdimensi tinggi. Dengan mengombinasikan BERT dan *Random Forest*.

Pendekatan yang digunakan mengombinasikan BERT sebagai metode untuk mengekstraksi makna dan topik dari judul skripsi mahasiswa dengan *Random Forest* sebagai algoritma klasifikasi. BERT berperan dalam memahami konteks

judul skripsi secara menyeluruh sehingga informasi topik dapat direpresentasikan secara lebih akurat. Selanjutnya, hasil representasi tersebut diolah menggunakan Random Forest untuk mempelajari pola kesesuaian antara topik skripsi dan bidang keahlian dosen pembimbing. Dengan menggabungkan kemampuan pemahaman semantik dari BERT dan kekuatan klasifikasi Random Forest, sistem rekomendasi yang dikembangkan diharapkan mampu menghasilkan rekomendasi dosen pembimbing yang lebih objektif, akurat, dan efisien dibandingkan metode konvensional.

Berdasarkan uraian latar belakang, penentuan dosen pembimbing skripsi yang masih mengandalkan proses manual dan pencocokan kata belum mampu merepresentasikan makna topik penelitian secara utuh. Dengan demikian, diperlukan suatu pendekatan yang mampu memahami konteks judul skripsi secara lebih mendalam guna meningkatkan kesesuaian antara topik penelitian mahasiswa dan bidang keahlian dosen pembimbing dalam proses penentuan pembimbing skripsi.

1.2. Rumusan Masalah

Berdasarkan permasalahan yang telah dijelaskan pada bagian latar belakang, penelitian ini difokuskan pada penerapan model BERT dalam mengidentifikasi topik judul skripsi mahasiswa serta pemanfaatan algoritma Random Forest untuk menentukan rekomendasi dosen pembimbing yang sesuai dengan topik penelitian tersebut.

1.3. Batasan Masalah

Penelitian ini memiliki batasan permasalahan yang dibahas agar dapat menyelesaikan permasalahan utama. Berikut ini batasan masalah penelitian ini :

1. Penelitian ini menggunakan data berupa judul skripsi mahasiswa serta data dosen pembimbing yang bersumber dari repository dan data akademik pada Universitas Muhammadiyah Sumatera Utara.
2. Pengolahan teks judul skripsi difokuskan pada penerapan model BERT untuk memperoleh representasi makna dan topik penelitian.
3. Penentuan rekomendasi dosen pembimbing dilakukan dengan memanfaatkan algoritma Random Forest berdasarkan hasil representasi teks dari BERT.
4. Penelitian ini tidak membahas penggunaan metode pemrosesan bahasa alami atau algoritma klasifikasi lain di luar BERT dan Random Forest.
5. Analisis hanya dilakukan pada teks judul skripsi tanpa melibatkan abstrak maupun isi skripsi secara keseluruhan.

1.4. Tujuan Penelitian

Tujuan umum dari penelitian ini adalah merancang dan membangun sebuah Sistem Rekomendasi Dosen Pembimbing yang cerdas dan efisien. Adapun tujuan spesifik penelitian ini didasarkan langsung pada rumusan masalah yang telah ditetapkan, yaitu:

1. Menerapkan model BERT untuk dapat mengenali dan mengekstrak topik skripsi dari data judul mahasiswa secara efektif.
2. Menerapkan algoritma *Random Forest* untuk membangun model yang mampu membuat keputusan rekomendasi dosen pembimbing berdasarkan hasil pengenalan topik oleh BERT.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan, baik secara teoritis maupun praktis:

A. Manfaat Teoritis

1. Pengembangan Model Hibrida: Membuktikan efektivitas penggabungan model bahasa canggih (BERT) sebagai pengestrak makna semantik dengan algoritma klasifikasi kuat (Random Forest) dalam tugas sistem rekomendasi berbasis konten.
2. Validasi NLP Bahasa Indonesia: Menyediakan referensi dan validasi metodologi *Natural Language Processing* (NLP) mutakhir (seperti BERT) ketika diterapkan pada data akademik yang spesifik (judul skripsi) dalam konteks Bahasa Indonesia.

B. Manfaat Praktis

Penelitian ini diharapkan dapat membantu program studi dalam menentukan dosen pembimbing skripsi secara lebih objektif serta mengurangi ketidaksesuaian antara topik penelitian mahasiswa dan bidang keahlian dosen.

BAB II

LANDASAN TEORI

2.1 Machine Learning

Machine learning merupakan cabang dari kecerdasan buatan yang bertujuan untuk melatih mesin menggunakan contoh atau dataset sesuai dengan tugas yang akan dilakukan. Melalui proses tersebut, mesin mempelajari pola dari data dan membentuk aturan sendiri, sehingga mampu mengenali serta mengolah data baru dengan lebih akurat. Penerapan Machine Learning sangat luas, mencakup berbagai bidang seperti pengenalan suara dan citra, analisis data, sistem rekomendasi, deteksi penipuan, pengenalan pola, serta pemrosesan bahasa alami. Seiring dengan pesatnya perkembangan teknologi dan meningkatnya ketersediaan data, Machine Learning terus berkembang dan menjadi salah satu bidang paling penting dalam ilmu komputer modern (Ananda, 2023).

Machine learning adalah sistem yang bisa belajar sendiri untuk mengambil keputusan tanpa harus selalu diprogram oleh manusia. Dengan cara ini, komputer bisa menjadi lebih “pintar” karena belajar dari data yang ada. Berdasarkan cara belajarnya, ada dua jenis utama: *supervised learning*, yaitu pembelajaran dengan data yang sudah punya label, dan *unsupervised learning*, yaitu pembelajaran yang mencari pola dari data tanpa label. Dalam penerapannya, dataset digunakan agar mesin bisa menghasilkan analisis yang lebih akurat.

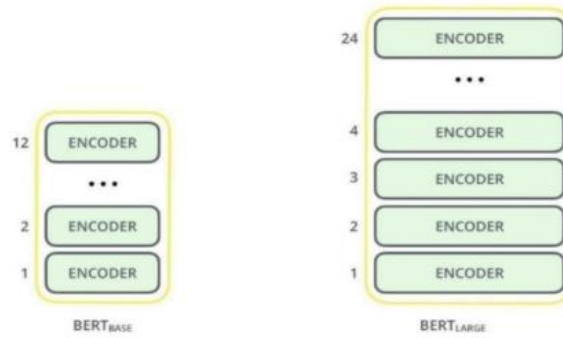
Salah satu fungsi utama machine learning adalah membuat komputer mampu belajar dari pengalaman agar dapat mengambil keputusan dengan lebih cerdas. Melalui proses ini, komputer dilatih untuk mengenali pola data dan mengklasifikasikannya ke dalam kelompok tertentu, kemudian mampu

mengevaluasi data baru yang belum dikenal. Klasifikasi, yang termasuk dalam *unsupervised learning*, membutuhkan persiapan data yang baik serta pemilihan bahasa pemrograman yang tepat agar proses pembelajaran mesin berjalan efektif dan menghasilkan prediksi yang akurat.

Dalam penerapannya, algoritma *machine learning* memungkinkan perguruan tinggi melakukan deteksi dini terhadap mahasiswa yang berisiko gagal akademik. Dengan menganalisis pola dari data historis, dosen atau pihak akademik dapat memberikan dukungan tambahan bagi mahasiswa yang membutuhkan. Data dari *Learning Management System* (LMS), seperti hasil kuis dan aktivitas diskusi, juga dapat dimanfaatkan untuk meningkatkan kualitas pembelajaran berbasis data. Dengan penerapan yang tepat, *machine learning* dapat membantu institusi pendidikan meningkatkan efektivitas pembelajaran, memberikan intervensi dini, serta mendukung keberhasilan akademik dan kelulusan mahasiswa secara lebih optimal (Azis, 2024).

2.2 directional Encoder Representations from Transformers (BERT)

BERT (Bidirectional Encoder Representations from Transformers) dikembangkan oleh tim Google AI pada tahun 2018. Model ini merupakan salah satu algoritma *deep learning* yang digunakan untuk memahami bahasa manusia (*Natural Language Processing*). BERT bekerja dengan cara mengenali makna kata berdasarkan konteks di sekitarnya menggunakan arsitektur *Transformer*. Dengan pendekatan ini, BERT mampu memahami hubungan antar kata dalam kalimat secara lebih mendalam. Model ini memiliki dua versi utama, yaitu BERT-base dan BERT-large, yang perbandingannya ditunjukkan pada Gambar 2.1.



Gambar 2.1 Ukuran BERT_{BASE} dan BERT_{LARGE}

(Sumber : (Kenton et al., 2019))

Perbedaan antara BERT base dan BERT large terletak pada ukuran dan kompleksitas modelnya. Berikut adalah perbedaan utama antara keduanya:

1. Ukuran Model

- 1) BERT base: Model BERT base memiliki ukuran yang relatif lebih kecil dibandingkan dengan BERT large. Biasanya, BERT base memiliki sekitar 110 juta parameter.
- 2) BERT large: Model BERT large memiliki ukuran yang lebih besar dibandingkan dengan BERT base. Biasanya, BERT large memiliki sekitar 340 juta parameter.

2. Jumlah Layer

- 1) BERT base: Model BERT base terdiri dari 12 layer transformasi (transformer blocks) dalam encoder.
- 2) BERT large: Model BERT large terdiri dari 24 layer transformasi (transformer blocks) dalam encoder.

3. Kompleksitas dan Representasi

- 1) BERT Base sudah cukup kuat untuk berbagai tugas pemrosesan bahasa alami seperti analisis sentimen, klasifikasi teks, dan pemahaman bahasa.

- 2) BERT Large memiliki kemampuan yang lebih tinggi dalam memahami konteks dan hubungan antar kata, sehingga lebih cocok untuk tugas-tugas yang lebih rumit.

4. Waktu dan Sumber Daya

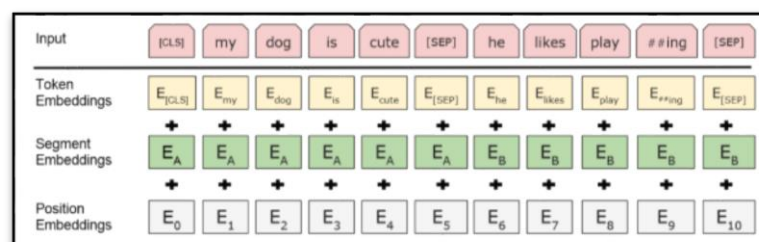
- 1) Karena ukurannya lebih besar, BERT Large memerlukan waktu pelatihan dan sumber daya komputasi yang lebih besar.
- 2) BERT Base lebih ringan dan efisien, sehingga lebih mudah digunakan untuk penelitian atau proyek dengan sumber daya terbatas.

Secara umum, pemilihan antara keduanya tergantung pada kebutuhan. Jika tugasnya sederhana, BERT Base sudah memadai. Namun, untuk tugas yang kompleks dan membutuhkan pemahaman konteks yang lebih dalam, BERT Large merupakan pilihan yang lebih baik.

BERT merupakan salah satu model yang paling populer dan efektif dalam pemrosesan bahasa alami karena kemampuannya memahami konteks kata secara mendalam melalui pendekatan dua arah. Model ini sangat unggul dalam menganalisis teks yang kompleks, seperti judul tugas akhir dan abstrak penelitian, karena mampu menangkap makna dan hubungan antar kata secara lebih menyeluruh dibandingkan metode lain yang hanya berfokus pada urutan kata. Dengan penerapan BERT, sistem rekomendasi diharapkan dapat memberikan hasil yang lebih akurat dan relevan dalam mencocokkan mahasiswa dengan dosen pembimbing berdasarkan kesesuaian topik penelitian, serta membantu meningkatkan efisiensi proses bimbingan dan kualitas penelitian mahasiswa (No et al., 2024).

Algoritma BERT adalah model yang digunakan untuk memahami arti kata dalam kalimat secara lebih mendalam. Berbeda dari metode lain, BERT membaca teks dari dua arah, yaitu dari kiri ke kanan dan dari kanan ke kiri, sehingga mampu memahami konteks dan makna setiap kata dengan lebih baik. BERT bekerja dengan menebak kata yang hilang dalam kalimat serta mengenali hubungan antara dua kalimat yang berbeda. Model ini sudah dilatih menggunakan banyak data teks sebelum digunakan untuk berbagai tugas, seperti analisis teks, klasifikasi data, dan pencarian informasi secara lebih akurat dan efisien. Dengan kemampuannya tersebut, BERT dapat membantu sistem komputer memahami bahasa manusia secara lebih alami, misalnya dalam mendeteksi topik, menganalisis sentimen, atau memberikan rekomendasi yang relevan berdasarkan isi teks (Hanum et al., 2024).

BERT memiliki cara unik dalam merepresentasikan teks secara berurutan. Pada awal setiap kalimat, ditambahkan token khusus [CLS] yang berfungsi untuk proses klasifikasi, sedangkan token [SEP] digunakan di akhir kalimat sebagai penanda batas antar kalimat. Selain itu, BERT juga menambahkan *segment embedding* untuk membedakan kata dari masing-masing kalimat dan *position embedding* untuk menunjukkan posisi setiap kata dalam urutan teks. Ilustrasi mengenai cara BERT merepresentasikan teks secara berurutan dapat dilihat pada gambar 2.2.

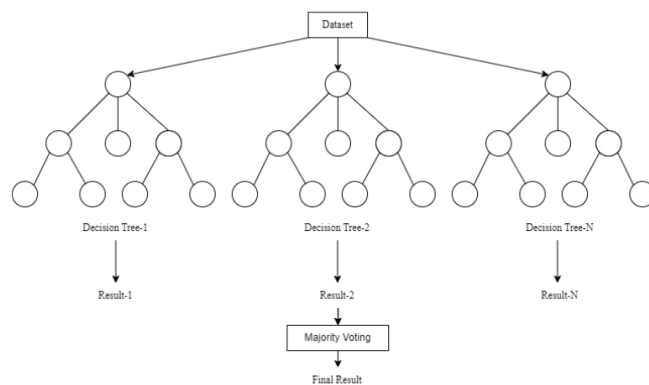


Gambar 2.2 Representasi Teks Sekuensial pada BERT

(Sumber : (Kenton et al., 2019))

2.3 Random Forest

Random Forest adalah algoritma yang banyak digunakan untuk tugas klasifikasi dalam *data mining*. Algoritma ini bekerja dengan menggabungkan hasil dari banyak pohon keputusan (*decision tree*) yang dibuat secara acak. Setiap pohon dibangun dari sebagian data dan fitur yang dipilih secara acak menggunakan teknik *bootstrapping*. Setelah semua pohon selesai dibuat, hasil prediksi dari setiap pohon digabungkan menggunakan metode *voting* (untuk klasifikasi) atau *averaging* (untuk regresi) agar hasil akhirnya lebih akurat dan stabil. Keunggulan *Random Forest* adalah kemampuannya mencegah *overfitting*, memproses data dalam jumlah besar dengan cepat, dan tetap menghasilkan prediksi yang baik pada data baru. Proses kerjanya dimulai dengan membagi data menjadi data latih dan data uji. Setiap pohon dilatih secara terpisah, kemudian hasilnya digabungkan untuk menghasilkan keputusan akhir. Tahapan ini membantu model mengenali pola dalam data dan membuat prediksi dengan tingkat ketepatan yang tinggi (Fauzi et al., 2025).



Gambar 2.3 Ilustrasi *Random Forest*

(Sumber : Muhammad et al., 2019)

Adapun langkah-langkah dalam pembentukan model *Random Forest* adalah sebagai berikut :

- 1) Pengambilan Sampel: Algoritma *Random Forest* mengambil sejumlah sampel dari dataset secara acak dengan penggantian, sehingga sebuah data dapat terpilih lebih dari satu kali.
- 2) Membangun Pohon Keputusan: Dari setiap sampel, *Random Forest* membangun pohon keputusan, yang dapat berupa pohon regresi untuk memprediksi nilai kontinu atau pohon klasifikasi untuk memprediksi kategori. Pada setiap titik pemisahan (*split*) dalam pohon, hanya sebagian fitur yang dipilih secara acak, sehingga setiap pohon memiliki struktur yang berbeda. Pohon-pohon tersebut tumbuh dari akar hingga daun menggunakan metode CART (*Classification and Regression Tree*), melakukan pemisahan secara berulang hingga mencapai hasil optimal di setiap cabang, tanpa dilakukan pemangkasan (*pruning*).
- 3) Penggabungan Hasil: Setelah seluruh pohon selesai dibangun, prediksi dari masing-masing pohon digabungkan untuk memperoleh hasil akhir. Pada kasus regresi, hasil prediksi akhir dihitung sebagai rata-rata dari semua pohon, sedangkan pada kasus klasifikasi, hasil akhir ditentukan melalui voting mayoritas dari seluruh pohon (Li et al., 2022).

2.4 Sistem Rekomendasi Dosen Pembimbing

Penentuan pembimbing akademik yang sesuai dengan topik tugas akhir sangat penting karena berpengaruh langsung terhadap kelancaran dan kualitas penyelesaian studi mahasiswa. Namun, di banyak perguruan tinggi proses pemilihan pembimbing masih dilakukan secara manual sehingga berpotensi memakan waktu dan menimbulkan ketidaktepatan. Oleh karena itu, penelitian ini mengusulkan sistem rekomendasi pembimbing akademik yang memanfaatkan data

publikasi dan riwayat bimbingan dosen sebagai dasar penentuan kesesuaian, sehingga diharapkan dapat membantu mahasiswa memilih pembimbing yang tepat dan mendukung kelulusan tepat waktu.

Penunjukan pembimbing akademik yang memiliki keahlian sesuai dengan topik tugas akhir merupakan hal penting, karena pembimbing berperan penting dalam membimbing mahasiswa menyelesaikan tugas akhir secara efektif. Kesesuaian antara keahlian pembimbing dan kebutuhan mahasiswa sangat memengaruhi keberhasilan studi. Di banyak perguruan tinggi, pemilihan pembimbing akademik masih dilakukan secara manual oleh koordinator tugas akhir, yang membutuhkan waktu lama dan berpotensi menimbulkan kesalahan manusia (Kurniawan, 2024).

Penelitian ini mengembangkan sistem rekomendasi pembimbing akademik yang memanfaatkan analisis publikasi dan riwayat bimbingan dosen dengan menggabungkan BERT dan Random Forest. Profil publikasi setiap dosen diolah menjadi satu kumpulan teks dan dianalisis menggunakan BERT untuk memperoleh representasi semantik topik penelitian. Hasil representasi tersebut kemudian digunakan oleh algoritma Random Forest untuk memprediksi tingkat kesesuaian antara topik tugas akhir mahasiswa dan keahlian dosen. Melalui pendekatan ini, sistem diharapkan mampu membantu mahasiswa memperoleh pembimbing yang relevan serta mendukung peningkatan efektivitas proses pembimbingan akademik dan ketepatan waktu penyelesaian studi.

2.5 Penelitian Terdahulu

Tabel 2.1 Ringkasan Penelitian Terdahulu

NO	Judul dan peneliti	Pembahasan	Metode	Kelebihan atau Kekurangan
1	Sistem Rekomendasi Dosen Pembimbing Skripsi Menggunakan Metode Cosine Similarity (Ali Akbar, Achmad Affandi & A. A. Hidayatullah 2021)	Penelitian ini membangun sistem rekomendasi untuk menentukan dosen pembimbing berdasarkan kesamaan antara topik skripsi mahasiswa dan bidang keahlian dosen menggunakan perhitungan <i>cosine similarity</i> .	<i>Cosine Similarity</i>	Kelebihan: Mudah di implementasikan, hasil cukup akurat untuk teks dengan topik spesifik. Kekurangan: Tidak menangkap makna semantik kata dan sangat bergantung pada kesamaan literal antar judul.
2	Thesis Supervisor Recommendation with	Penelitian ini mengembangkan sistem	<i>Information Retrieval</i> ,	Kelebihan: Meningkatkan akurasi

	Representative Content and Information Retrieval (Maresha Caroline Wijanto, Rachmi Rachmadiany & Oscar Karnalim 2020)	rekomendasi dosen pembimbing berdasarkan kesamaan isi antara proposal skripsi mahasiswa dan publikasi akademik dosen. Sistem menilai relevansi dengan menghitung kemiripan teks menggunakan <i>vector space model</i> dan <i>cosine similarity</i>	<i>Vector Space Model</i> , dan <i>Cosine Similarity</i> .	rekomendasi melalui analisis isi publikasi dan grouping keahlian dosen. Kekurangan: Kinerja bergantung pada kelengkapan data publikasi dan pembaruan profil dosen.
3	Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing Berdasarkan Judul Tugas Akhir (Ferris	Penelitian ini merancang sistem rekomendasi dosen pembimbing dengan mencocokkan	<i>BERT</i> dan <i>Cosine Similarity</i>	Kelebihan: Akurasi tinggi (90%) dan mampu memahami kesamaan semantik

	Tita Sabilillah, Sri Winarno & Ryandhika Bintang Abiyyi 2024)	judul tugas akhir mahasiswa terhadap penelitian dosen menggunakan <i>BERT</i> untuk representasi teks dan <i>Cosine Similarity</i> untuk menghitung tingkat kemiripan. Data diambil melalui <i>web scraping</i> dari Google Scholar dan ResearchGate.		antartajudul. Kekurangan: Ketergantungan pada kelengkapan data publikasi dosen serta belum mempertimbangan faktor non-akademik seperti preferensi mahasiswa.
4	Optimalisasi <i>Random Forest</i> untuk Sentimen Bahasa Indonesia dengan GridSearch dan SMOTE	Penelitian ini mengoptimalkan algoritma <i>Random Forest</i> untuk analisis sentimen media sosial X (Twitter)	Random Forest, SMOTE, GridSearch, TF-IDF, TextBlob	Kelebihan : Dapat mengatasi data tidak seimbang dengan SMOTE, Akurasi meningkat dari 73% - 89%

	(Ahmad Fauzi, Agus Heri Yunial, Dede Eko Saputro, Reza Saputra 2025)	berbahasa Indonesia dengan topik “UKT”. Labeling dilakukan otomatis dengan TextBlob, keseimbangan data diperbaiki menggunakan SMOTE, dan performa model ditingkatkan lewat Hyperparameter GridSearch.		setelah optimasi, Kombinasi metode menghasilkan model yang lebih stabil dan kuat Kelemahan : TextBlob perlu penerjemahan ke Bahasa Inggris (belum mendukung langsung Bahasa Indonesia)
5	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Jacob Devlin, Ming-Wei Chang,	BERT memperkenalkan model representasi bahasa bidirectional Transformer yang dilatih secara pre-	BERT (Bidirectional Encoder Representations from Transform	Kelebihannya mampu memahami konteks kalimat secara menyeluruh dan menghasilkan akurasi tinggi,

	Kenton Lee, Kristina Toutanova 2019)	training menggunakan dua tugas utama: <i>Masked Language Model (MLM)</i> dan <i>Next Sentence Prediction (NSP)</i> . Model ini mampu memahami konteks dari dua arah (kiri dan kanan) dan dapat disesuaikan (<i>fine-tuned</i>) untuk berbagai tugas NLP seperti klasifikasi, QA, dan NER.	ers) dengan dua tahap pelatihan: <i>Masked Language Model (MLM)</i> dan <i>Next Sentence Prediction (NSP)</i>	Kekurangannya membutuhkan sumber daya komputasi besar dan waktu pelatihan yang lama.
--	--	---	---	--

2.6 ANALISIS GAP

Berdasarkan hasil kajian dari beberapa penelitian sebelumnya, sistem rekomendasi dosen pembimbing sudah banyak dikembangkan dengan berbagai metode pembelajaran mesin. Penelitian yang dilakukan oleh Ali Akbar dkk. (2021) menggunakan *Cosine Similarity* untuk menghitung kesamaan antara judul skripsi mahasiswa dan bidang keahlian dosen. Metode ini mudah digunakan, tetapi hanya

melihat kesamaan kata tanpa memahami makna dari isi kalimat. Sementara itu, penelitian Wijanto dkk. (2020) menggunakan pendekatan *Information Retrieval* dan *Vector Space Model* untuk mencari kesamaan antara proposal mahasiswa dan publikasi dosen. Metode ini cukup meningkatkan akurasi, namun masih terbatas dalam memahami makna kalimat dan sangat bergantung pada kelengkapan data publikasi dosen.

Penelitian lain oleh Sabilillah dkk. (2024) menggabungkan BERT dan *Cosine Similarity* untuk menghasilkan sistem rekomendasi dosen pembimbing dengan hasil yang lebih akurat, karena BERT mampu memahami arti dan konteks antarjudul. Namun, penelitian ini belum menggunakan model klasifikasi tambahan yang dapat memperkuat hasil prediksi. Di sisi lain, penelitian Fauzi dkk. (2025) menunjukkan bahwa algoritma *Random Forest* yang dikombinasikan dengan *GridSearch* dan *SMOTE* dapat meningkatkan akurasi klasifikasi teks berbahasa Indonesia. Walaupun hasilnya baik, penelitian ini belum diterapkan dalam konteks rekomendasi akademik seperti penentuan dosen pembimbing.

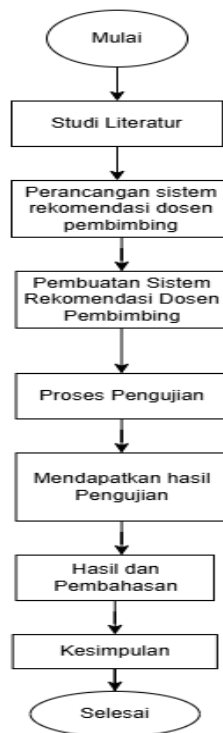
Berdasarkan penelitian sebelumnya, belum ditemukan kajian yang secara khusus menggabungkan BERT dan Random Forest untuk sistem rekomendasi dosen pembimbing dengan data berbahasa Indonesia. Oleh karena itu, penelitian ini menggabungkan kemampuan BERT dalam memahami makna teks dan Random Forest dalam klasifikasi guna meningkatkan akurasi rekomendasi dosen pembimbing berdasarkan judul skripsi mahasiswa.

BAB III

METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan dalam penelitian ini adalah penelitian eksperimental dengan pendekatan kuantitatif. Penelitian ini memanfaatkan model BERT untuk mengekstraksi fitur semantik dari judul skripsi serta algoritma Random Forest untuk melakukan klasifikasi dan rekomendasi dosen pembimbing berdasarkan tingkat kesesuaian topik.

3.1 Alur Penelitian

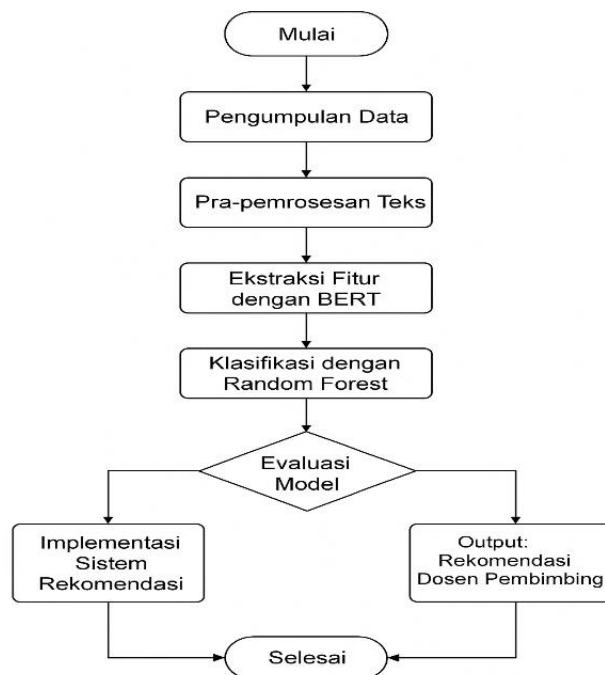


Gambar 3.1 Alur Penelitian

Gambar tersebut menggambarkan alur penelitian yang dimulai dari tahap studi literatur, yaitu mengumpulkan dan mempelajari berbagai teori serta penelitian sebelumnya yang berkaitan dengan sistem rekomendasi dosen pembimbing. Setelah itu, dilakukan perancangan sistem untuk menentukan bagaimana sistem

akan bekerja dan fitur apa saja yang diperlukan. Tahap berikutnya adalah pembuatan sistem rekomendasi dosen pembimbing, di mana rancangan yang telah dibuat diimplementasikan menjadi sebuah sistem yang dapat dijalankan. Setelah sistem selesai dibuat, dilakukan proses pengujian untuk memastikan bahwa sistem berfungsi dengan baik dan menghasilkan rekomendasi yang sesuai. Hasil dari pengujian tersebut kemudian dianalisis dalam tahap hasil dan pembahasan guna mengetahui tingkat akurasi dan keefektifan sistem. Berdasarkan hasil analisis tersebut, peneliti kemudian menyusun kesimpulan yang merangkum temuan utama dari penelitian. Setelah seluruh tahap selesai dilaksanakan, penelitian ini dinyatakan selesai.

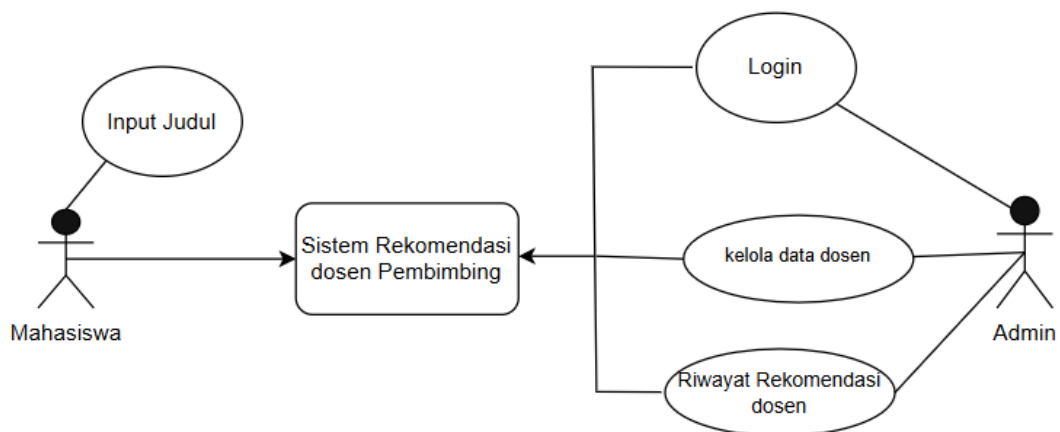
3.2 Diagram Rancangan Sistem dan Pemodelan UML



Gambar 3.2 Diagram Rancangan Sistem

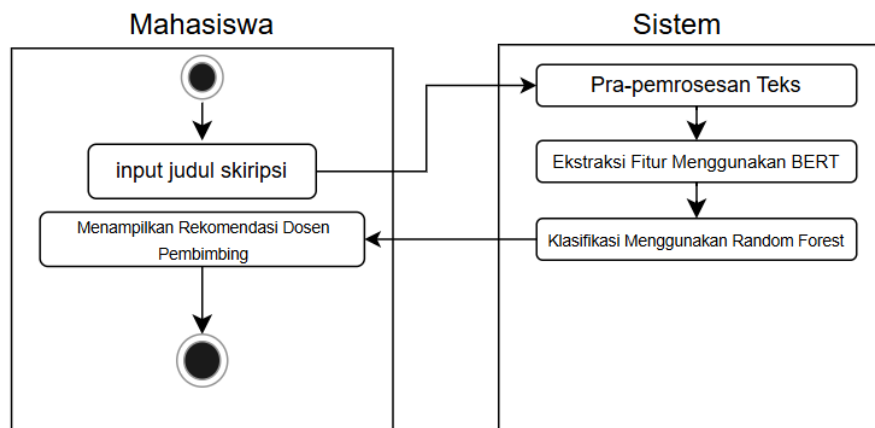
Sistem ini merupakan sistem rekomendasi dosen pembimbing skripsi yang dibangun menggunakan BERT dan Random Forest. Proses kerja sistem dimulai

dengan pengumpulan data berupa judul skripsi mahasiswa dan profil dosen. Data tersebut kemudian melalui tahap pra-pemrosesan teks untuk membersihkan dan menyeragamkan format. Selanjutnya, sistem melakukan ekstraksi fitur menggunakan BERT guna memperoleh representasi makna dari teks secara mendalam. Hasil ekstraksi tersebut diklasifikasikan menggunakan *Random Forest* untuk mencocokkan topik skripsi dengan bidang keahlian dosen. Setelah itu, sistem melakukan evaluasi model untuk menilai kinerja klasifikasi. Jika hasilnya memenuhi kriteria, sistem menghasilkan rekomendasi dosen pembimbing yang paling sesuai dengan topik penelitian mahasiswa.



Gambar 3.3 Use Case Diagram

Pada gambar diatas Use Case Diagram interaksi antara pengguna dan sistem rekomendasi pembimbing akademik. Mahasiswa berperan untuk menginput judul skripsi dan melihat hasil rekomendasi, sedangkan admin bertugas untuk login, mengelola data yang digunakan sistem seperti data dosen dan riwayat rekomendasi dosen pembimbing.



Gambar 3.4 Activity Diagram

Gambar diatas Activity Diagram menunjukkan alur aktivitas sistem rekomendasi pembimbing akademik yang dimulai dari input judul skripsi hingga sistem menampilkan rekomendasi dosen pembimbing.

3.3 Sumber dan Jenis data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari *repository* skripsi mahasiswa Universitas Muhammadiyah Sumatera Utara terutama fakultas Ilmu Komputer dan teknologi informasi serta database dosen pada lingkungan universitas yang dapat dari Google Scholar. Data tersebut digunakan untuk membangun dan melatih model sistem rekomendasi dosen pembimbing berdasarkan kesesuaian topik dan kata kunci penelitian.

Data yang dikumpulkan terdiri atas dua komponen utama, yaitu:

- 1) judul skripsi dan topik penelitian yang diajukan oleh mahasiswa.
- 2) Data dosen, yang mencakup nama dosen, bidang keahlian, serta kata kunci penelitian yang menggambarkan area kompetensinya.

Seluruh data kemudian diproses dan diubah menjadi dataset teks yang siap digunakan dalam proses pelatihan (*training data*) dan pengujian (*testing data*) model rekomendasi. Dataset ini berfungsi sebagai dasar bagi sistem dalam melakukan analisis semantik menggunakan BERT dan klasifikasi dengan *Random Forest* untuk menghasilkan rekomendasi dosen pembimbing yang relevan.

3.4 Variabel Penelitian

Variabel penelitian merupakan komponen penting yang digunakan untuk mengukur, menganalisis, dan mengembangkan model yang diusulkan. Dalam penelitian ini, variabel dibagi menjadi dua jenis, yaitu variabel input dan variabel output.

1. Variabel input

Judul skripsi berfungsi sebagai sumber utama untuk menganalisis konteks dan makna semantik menggunakan model BERT, sehingga sistem dapat memahami bidang penelitian yang ingin dikaji oleh mahasiswa. Representasi teks dari judul skripsi kemudian diubah menjadi vektor numerik (*embedding*) yang akan digunakan sebagai masukan bagi algoritma klasifikasi.

2. Variabel output

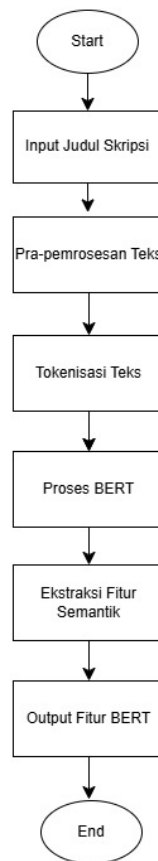
Variabel ini berupa dosen pembimbing yang direkomendasikan berdasarkan tingkat kesesuaian topik dan bidang keahlian dosen. Hasil keluaran diperoleh melalui proses klasifikasi menggunakan Random Forest, yang memprediksi dosen paling relevan terhadap topik skripsi mahasiswa. Output ini menjadi hasil akhir dari sistem rekomendasi yang dirancang

untuk membantu proses penentuan pembimbing akademik secara lebih objektif dan efisien.

3.5 Tahap Ekstraksi Fitur Menggunakan BERT

Tahap ekstraksi fitur menggunakan BERT (Bidirectional Encoder Representations from Transformers) bertujuan untuk mengubah teks menjadi representasi numerik (*embedding*) yang dapat dipahami oleh algoritma *machine learning*. BERT merupakan model bahasa berbasis *transformer* yang dilatih secara dua arah (*bidirectional*), sehingga mampu memahami konteks kata dalam kalimat secara menyeluruh, baik dari arah kiri maupun kanan. Dengan kemampuan ini, BERT menghasilkan representasi semantik yang lebih akurat dibandingkan metode tradisional seperti *TF-IDF* yang hanya memperhatikan frekuensi kemunculan kata tanpa mempertimbangkan makna konteksnya.

Pada tahap ini, teks hasil pra-pemrosesan berupa judul skripsi mahasiswa dan bidang keahlian dosen dimasukkan ke dalam model BERT. Setiap kata dalam kalimat diubah menjadi vektor berdimensi tinggi yang mencerminkan makna kata berdasarkan konteks kalimat secara keseluruhan. Proses ini melibatkan tokenisasi, *embedding* posisi, dan *encoding* menggunakan lapisan *transformer* yang memungkinkan BERT untuk memahami hubungan semantik antar kata dalam teks akademik.



Gambar 3.5 Flowchart BERT

Flowchart tersebut menggambarkan alur proses ekstraksi fitur menggunakan BERT yang dimulai dari input judul skripsi. Judul skripsi terlebih dahulu melalui tahap pra-pemrosesan dan tokenisasi teks, kemudian diproses menggunakan model BERT untuk memahami konteks dan makna teks. Hasil dari proses ini berupa fitur semantik yang merepresentasikan topik judul skripsi, yang selanjutnya digunakan sebagai output untuk tahap berikutnya.

Untuk memperjelas proses ekstraksi fitur menggunakan BERT, berikut contoh ilustrasi tahapan ekstraksi fitur dari judul skripsi mahasiswa. Contoh Ilustrasi, terdapat judul skripsi mahasiswa “Sistem Rekomendasi Dosen Pembimbing Skripsi Menggunakan BERT dan Random Forest” Judul tersebut

pertama kali diproses melalui tahap tokenisasi oleh tokenizer BERT, sehingga diubah menjadi token-token sebagai berikut: [CLS] Sistem Rekomendasi Dosen Pembimbing Skripsi Menggunakan BERT Dan Random Forest [SEP].

Setiap token kemudian dipetakan ke dalam vektor embedding berdimensi tetap (misalnya 768 dimensi pada BERT-base). Selanjutnya, seluruh token diproses oleh lapisan transformer BERT untuk menghasilkan representasi kontekstual. Representasi vektor dari token khusus [CLS] digunakan sebagai representasi keseluruhan judul skripsi, karena token ini merepresentasikan informasi global dari kalimat. Sebagai ilustrasi sederhana, hasil ekstraksi fitur BERT dapat digambarkan seperti pada Tabel berikut.

Tabel. 3.1 Representasi Fitur (Embedding BERT)

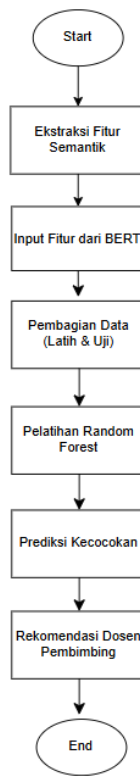
Judul Skripsi	Representasi Fitur (Embedding BERT)
sistem Rekomendasi Dosen Pembimbing Menggunakan BERT dan Random Forest	[0.021, -0.134, 0.567, ... , 0.089]
Pemodelan Sistem Rekomendasi Akademik Menggunakan Natural Language Processing	[-0.028, 0.245, 0.362, ... , -0.083]
Implementasi Machine Learning dalam Penentuan Kesesuaian Dosen Pembimbing Skripsi	[-0.061, 0.301, 0.287, ... , -0.049]
Analisis Sentimen Ulasan Pelanggan Terhadap Layanan E-Commerce Menggunakan BERT	[0.084, -0.219, 0.467, ... , 0.052]

3.6 Tahap Klasifikasi Menggunakan Random Forest

Tahap klasifikasi menggunakan Random Forest dilakukan setelah proses ekstraksi fitur menggunakan BERT selesai. Pada tahap ini, algoritma Random Forest tidak lagi memproses data teks, melainkan menerima vektor embedding hasil ekstraksi BERT sebagai input. Setiap judul skripsi mahasiswa direpresentasikan

dalam bentuk vektor numerik berdimensi tinggi yang mencerminkan makna dan topik penelitian. Vektor-vektor ini kemudian digunakan sebagai fitur untuk melatih model Random Forest dalam mempelajari pola kesesuaian antara topik skripsi mahasiswa dan bidang keahlian dosen pembimbing.

Model Random Forest dibangun dari sejumlah pohon keputusan yang masing-masing dilatih menggunakan subset data embedding BERT. Setiap pohon akan menghasilkan prediksi dosen pembimbing yang sesuai, kemudian Hasil akhir rekomendasi diperoleh dengan menggabungkan hasil prediksi dari seluruh pohon keputusan yang ada.



Gambar 3.6 Flowchart *Random Forest*

Flowchart tersebut menunjukkan proses rekomendasi dosen pembimbing menggunakan algoritma Random Forest. Proses dimulai dengan ekstraksi fitur semantic lalu memasukkan fitur hasil ekstraksi dari BERT, kemudian data dibagi

menjadi data latih dan data uji. Selanjutnya, model Random Forest dilatih untuk mempelajari pola kecocokan, sehingga dapat menghasilkan prediksi kesesuaian dan memberikan rekomendasi dosen pembimbing yang paling relevan.

Dengan pendekatan ini, BERT berperan sebagai pengekstraksi fitur semantik, sedangkan Random Forest berperan sebagai pengklasifikasi yang menentukan dosen pembimbing berdasarkan fitur tersebut. Kombinasi ini memungkinkan sistem untuk tidak hanya memahami makna judul skripsi, tetapi juga menghasilkan keputusan rekomendasi yang lebih stabil dan akurat.

3.7 Rancangan Desain Interface Sistem

The image shows a web interface for an admin login. At the top, there is a header bar with the text 'Sistem Rekomendasi Dosen Pembimbing' on the left and '(LOG)' on the right. Below the header, the main content area contains a login form. The form has two input fields: one labeled 'Username' and one labeled 'Password'. Below these fields is a button labeled 'Login (DSA)'. The entire interface is enclosed in a rectangular border.

Gambar 3.7 Desain Interface Menu Login Admin

Tampilan Halaman ini (LOG) digunakan oleh admin untuk masuk ke dalam sistem dengan memasukkan *username* dan *password*. Proses login berfungsi sebagai mekanisme autentikasi agar hanya pengguna yang memiliki hak akses yang dapat menggunakan sistem. Dan Saat admin Masuk Mengklik login maka akan masuk ke halaman Dashboard admin (DSA).

Sistem Rekomendasi Dosen Pembimbing (DSA)			
Kelola Dosen	Nama Dosen	Keahlian	Aksi
(RRD) Riwayat Rekomendasi	Al-Khowarizmi	Kecerdasan Buatan	Edit Hapus
(LOG) Logout	Firahmi rizky	Sistem Informasi	Edit Hapus
	MHD.Basri	Internet Of things	Edit Hapus
	Fatma Sari H	Kecerdasan Buatan	Edit Hapus
	Mahardika	Ilmu Komputer	Edit Hapus

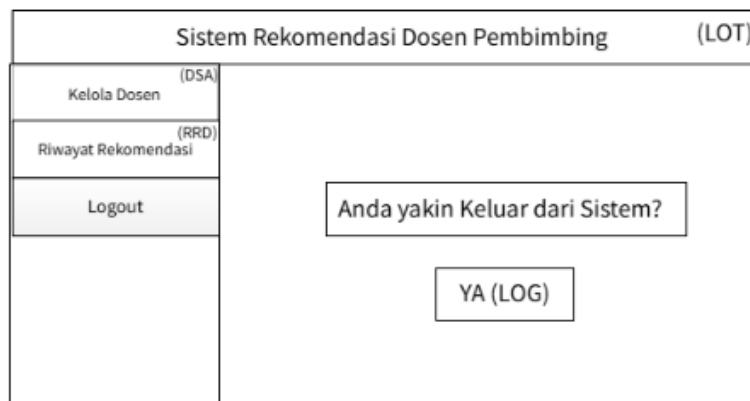
Gambar 3.8 Desain Interface Menu Utama Admin

Tampilan halaman pada gambar diatas merupakan antarmuka admin pada Sistem Rekomendasi Dosen Pembimbing. Admin memiliki hak akses untuk mengelola data dosen dan bidang keahlian melalui fitur kelola dosen (DSA). Antarmuka ini hanya dapat diakses setelah proses login , sehingga membedakan peran admin dengan mahasiswa yang hanya dapat mengakses fitur rekomendasi tanpa login.

Sistem Rekomendasi Dosen Pembimbing (RRD)		
(DSA) Kelola Dosen	Judul Skripsi	Dosen Pembimbing
Riwayat Rekomendasi	RANCANG BANGUN SISTEM	Al-Khowarizmi
(LOT) Logout	CLUSTERING JENIS SAYURAN	Firahmi Rizky
	PREDIKSI PENJUALAN SUSU	Fatma Sari H
	SISTEM MONITORING	MHD.Basri
	RANCANG BANGUN SISTEM	Mahardika

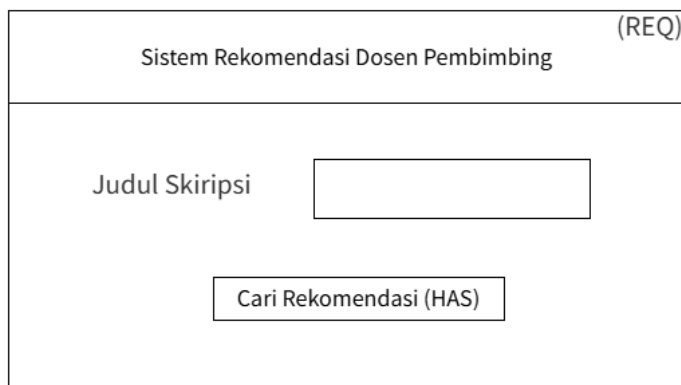
Gambar 3.9 Desain Interface Riwayat Rekomendasi

Tampilan halaman pada gambar diatas merupakan antarmuka admin pada Sistem Rekomendasi Dosen Pembimbing pada halaman Riwayat Rekomendasi. Admin memiliki hak akses untuk mengelola Judul Skripsi dan nama dosen pembimbing (RRD).



Gambar 3.10 Desain Interface menu Logout

Tampilan halaman pada gambar diatas merupakan antarmuka admin pada Sistem Rekomendasi Dosen Pembimbing pada halaman Logout (LOT). Dimana admin bisa keluar dari sistem dengan mengklik Ya.



Gambar 3.11 Desain Halaman Input Judul Skripsi Mahasiswa

Tampilan tersebut menunjukkan antarmuka mahasiswa pada Sistem Rekomendasi Dosen Pembimbing (REQ). Pada halaman ini, mahasiswa dapat memasukkan judul skripsi yang akan diajukan. Selanjutnya, mahasiswa menekan tombol “Cari Rekomendasi” untuk memulai proses analisis judul skripsi (HAS). Judul skripsi yang diinputkan akan diproses menggunakan metode BERT untuk mengekstraksi fitur semantik, kemudian diklasifikasikan menggunakan algoritma Random Forest untuk menghasilkan rekomendasi dosen pembimbing yang sesuai

dengan topik penelitian. Antarmuka ini dirancang tanpa proses login guna mempermudah akses mahasiswa serta mempertahankan fokus penelitian pada kinerja sistem rekomendasi.

Sistem Rekomendasi Dosen Pembimbing (HAS)	
Judul	<input type="text"/>
Nama Dosen	<input type="text"/>
Keahlian	<input type="text"/>
	<input type="button" value="OK (REQ)"/>

Gambar 3.12 Desain Hasil Rekomendasi Dosen Pembimbing

Tampilan ini menunjukkan Hasil Rekomendasi Dosen Pembimbing yang berisi informasi Judul Skripsi, Nama Dosen, dan Keahlian (HAS). Data ditampilkan sebagai hasil rekomendasi sistem, dan pengguna dapat menekan tombol ok untuk mengonfirmasi dan kembali pada halaman (REQ).

3.8 Jadwal Penelitian

Tabel 3.2 Jadwal Penelitian

No	Nama Kegiatan	Bulan Ke																											
		November				Desember				Januari				Februari				Maret				Mei				Juni			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengajuan Judul																												
2	Penulisan Bab I - Bab III																												
3	Pengumpulan data																												
4	Pembuatan sistem																												
5	Pengujian Sistem																												
6	Penulisan Bab IV - Bab V																												

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

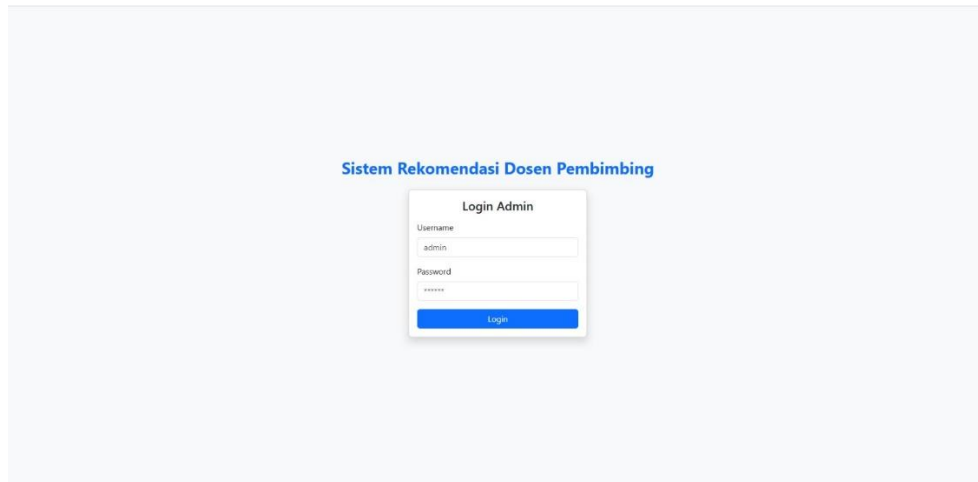
Hasil tampilan antar muka merupakan tahapan dimana sistem sudah siap untuk dioperasikan pada kondisi sebenarnya sesuai hasil analisis dan perancangan yang telah dilakukan. Melalui tahapan ini dapat diketahui apakah sistem yang dibangun mampu mencapai tujuan yaitu memberikan rekomendasi dosen pembimbing yang sesuai berdasarkan judul skripsi yang diinputkan. Aplikasi ini dilengkapi dengan tampilan antarmuka yang dirancang untuk memudahkan pengguna dalam proses penginputan judul dan melihat hasil rekomendasi dosen. Fungsi dari interface (antarmuka) ini adalah sebagai media untuk memberikan input berupa judul skripsi dan menampilkan output berupa nama dosen yang direkomendasikan beserta bidang keahliannya. Pada aplikasi ini antarmuka yang disediakan terdiri dari menu beranda, menu input judul skripsi, serta menu hasil rekomendasi.

4.1.1 Halaman Admin

Dalam Halaman Admin ada beberapa halaman akses. Adapun Halaman Admin sebagai berikut.

1. Menu Login

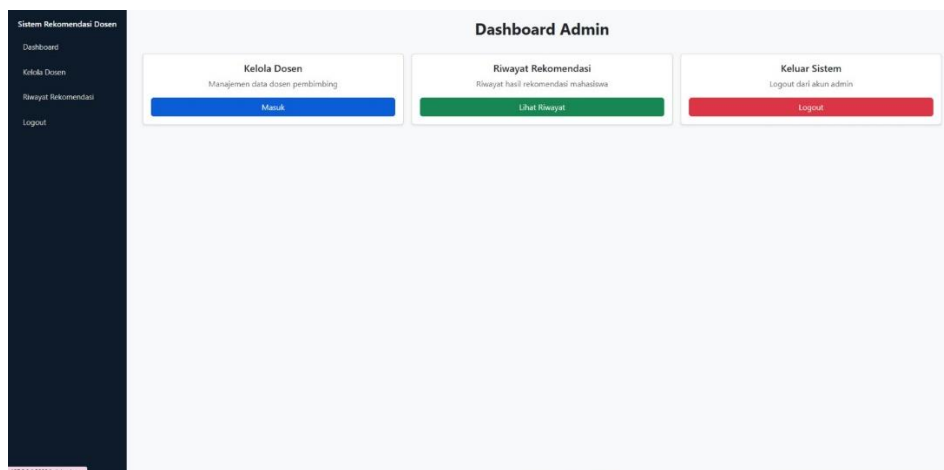
Menu Login digunakan untuk mengamankan sistem dari user-user yang tidak bertanggung jawab sebelum masuk ke menu utama. Berikut tampilannya :



Gambar 4.1 Menu Login

2. Menu Utama

Menu Utama digunakan Sebagai halaman pusat navigasi yang menghubungkan pengguna (admin) ke fitur-fitur utama sistem, seperti mengelola data dosen, melihat riwayat rekomendasi, serta mengakses fungsi lain dalam proses rekomendasi dosen pembimbing. Adapaun tampilannya sebagai berikut.



Gambar 4.2 Menu Utama

No	Nama Dosen	Keahlian	Aksi
1	Dr.Firaheni Rizky,S.Kom.,M.Kom	Artificial Intelligence, Expert system	Edit Hapus
2	Indah Purnama Sari,S.T.,M.Kom	Computer Science, Webs, Data Mining, Artificial Intelligence	Edit Hapus
3	Amrullah,S.Kom.,M.Kom	Artificial Intellegency, Machine Learning, Cryptography, Information System, Deep learning	Edit Hapus
4	Fatma Sari Hutagalung S.Kom.,M.Kom	Kecerdasan Buatan, Sistem Pendukung Keputusan, Data Base	Edit Hapus
5	Farid Akbar Siregar S.Kom.,M.Kom	Teknologi Informasi, Kriptografi, Decision Support System	Edit Hapus
6	Mahardika Abdi Prawira Tanjung S.Kom.,M.Kom	Computer Vision, Machine Learning, Educational Games, Data Science	Edit Hapus
7	Rizaldy Khair S.Kom.,M.Kom	Ilmu Komputer, Sistem Informasi, IOT, Software Engineering, Financial Technology	Edit Hapus
8	Ferdy Rizza,S.T.,M.Kom	Cryptography Security System	Edit Hapus
9	Yohannis Syahra,Si,M.Kom	Data Mining, Sistem Pakar	Edit Hapus
10	Yoshida Sari,S.E.,S.Kom.,M.Kom	Sistem Informasi, Teknologi Informatika	Edit Hapus
11	Zuli Agudina Galton,M.Si	Statistika, Data Science, Matematika	Edit Hapus
12	Mhd.Berli,Si,M.Kom	Basia Data, Query, Sistem Informasi, Web Application, Mikrokontroler dan IoT	Edit Hapus
13	Marliano,S.Pd.,S.Kom.,M.Kom	Data Mining, Information, System, Audit System	Edit Hapus
14	DR.Irfan,M.Si	Akuntansi, Akuntansi Zakat, Akuntansi Zakat Korporat, Metodologi Penelitian	Edit Hapus
15	Mulkan Azhari,S.Kom.,M.Kom	Ilmu komputer, Teknologi Informasi, Sistem Informasi, Data Sains	Edit Hapus
16	Halim Maulana,S.T.,M.Kom	Security System, Cloud Computing, Management of IT Computer Forensics, Data Science	Edit Hapus
17	Dr. M. Muhammad S. F. Kom., M. Kom.	Computer Science, Information, Technologies	Edit Hapus

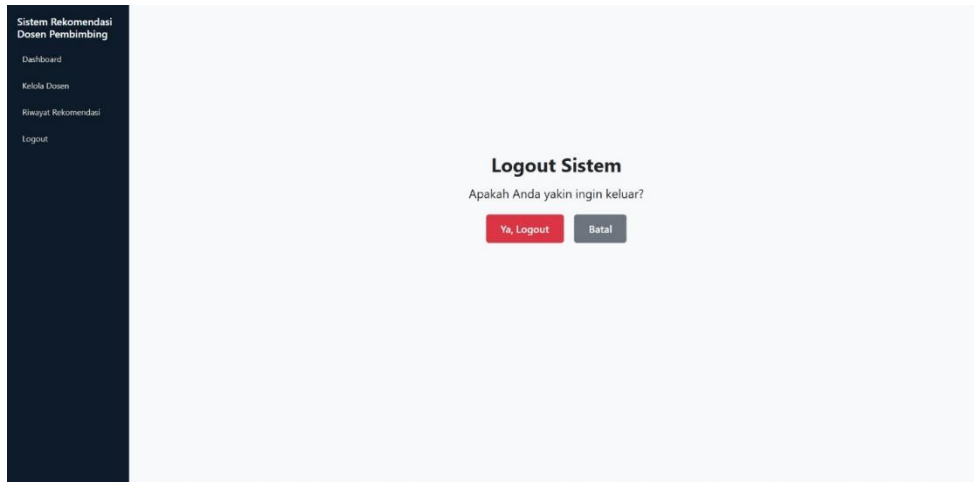
Gambar 4.3 Halaman Kelola Dosen

Pada halaman kelola dosen terdapat nama dosen dan keahliannya serta fitur untuk mengedit dan juga menghapus data dosen.

No	Judul Skripsi	Dosen Pembimbing	Aksi
1	CLUSTERING JENIS SAYURAN DI DAERAH DESA SEMPLAJAWA DENGAN ALGORITMA GAUSSIAN MIXTURE MODEL	Dr.Firaheni Rizky,S.Kom.,M.Kom	Edit Hapus
2	SISTEM CERDAS PENGENDALIAN LAMPU BERBASIS IOT DENGAN LOGIKA FUZZY UNTUK EFISIENSI ENERGI	Indah Purnama Sari,S.T.,M.Kom	Edit Hapus
3	MONITORING ALAT HITUNG TETAPAN INFRUS DENGAN ESP32	Amrullah,S.Kom.,M.Kom	Edit Hapus
4	PREDIKSI PENJUALAN SUSU LANGKA DENGAN METODE WEIGHTED MOVING AVERAGE DI TOKO SUSU ASIA KING	Fatma Sari Hutagalung S.Kom.,M.Kom	Edit Hapus
5	PENGOLAHAN CITRA DIGITAL UNTUK PENGENALAN JENIS DALIN SEHAT BERBASIS FITUR WARNA DAN BENTUK MENGGUNAKAN ALGORITMA K-NEARE	Farid Akbar Siregar S.Kom.,M.Kom	Edit Hapus
6	PERBANDINGAN ALGORITMA K-MEANS CLUSTERING DAN FUZZY TSUKAMOTO UNTUK MENGELOMPOKKAN DATA SISWA SMP NU MEDAN BERDASARKAN PRESTASI AKADEMIK	Amrullah,S.Kom.,M.Kom	Edit Hapus
7	SISTEM IRRIGASI CERDAS UNTUK TANAMAN GREENHOUSE MENGGUNAKAN LOGIKA FUZZY MAMDANI DAN PREDIKSI CUACA BERBASIS IOT	Mahardika Abdi Prawira Tanjung S.Kom.,M.Kom	Edit Hapus
8	PREDIKSI PERMINTAAN LAYANAN GORIDE DI KOTA MEDAN MENGGUNAKAN ALGORITMA LINEAR REGRESSION DAN VISUALISASI PETA INTERAKTIF PADA DATA PERUSAHAAN GOJEK	Rizaldy Khair S.Kom.,M.Kom	Edit Hapus
9	PENERAPAN METODE COSINE SIMILARITY PADA SOAL TES PENERIMAAN KARIRIWAN PT. DANA PURNA INVESTAMA	Indah Purnama Sari,S.T.,M.Kom	Edit Hapus
10	SISTEM KONTROL PERGANTIAN AIR PADA KOLAM BUDI DAYA IKAN LEELE BERBASIS INTERNET OF THINGS	Ferdy Rizza,S.T.,M.Kom	Edit Hapus

Gambar 4.4 Halaman Riwayat Rekomendasi

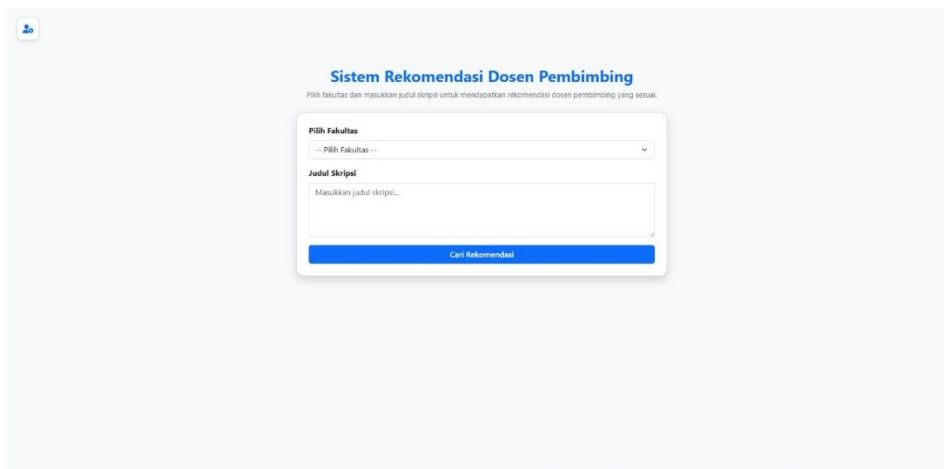
Pada halaman Riwayat Rekomendasi terdapat judul skripsi dan nama dosen pembimbingnya yang didapatkan dari repository Universitas Muhammadiyah Sumatera Utara serta fitur edit dan hapus judul Skripsi.



Gambar 4.5 Halaman Logout

Pada halaman logout terdapat pilihan yaitu keluar dari sistem atau batal keluar sistem, yang mana jika dipilih “ya, Logout” maka akan kembali pada menu login, jika batal maka akan tetap pada sistem.

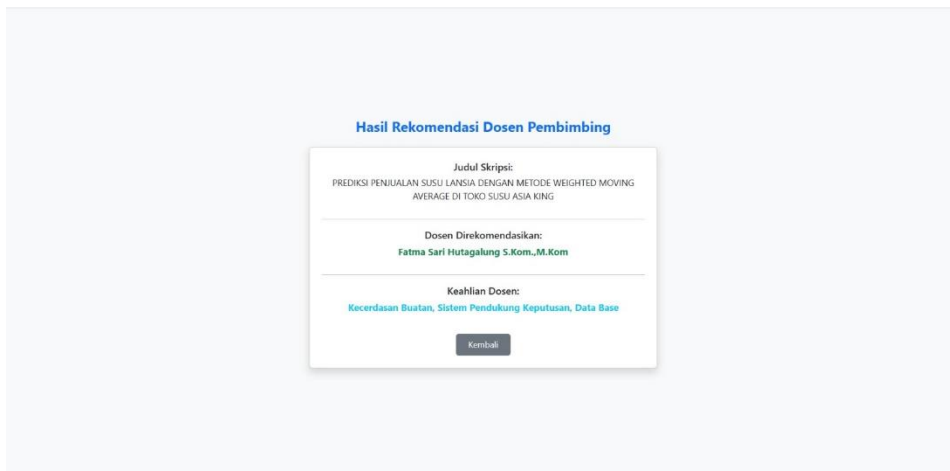
4.1.2 Halaman Input Judul



Gambar 4.6 Halaman input judul

Pada halaman ini terdapat kolom untuk memilih fakultas dan memasukkan judul skripsi dan tombol untuk mencari rekomendasi dosen pembimbing, ini bisa diakses siapa saja untuk melihat hasil rekomendasi dosen pembimbing.

4.1.3 Hasil Input



Gambar 4.7 Halaman Hasil Input

Pada halaman ini terdapat judul skripsi yang di input, beserta hasil dosen rekomendasi dan juga bidang keahlian dosen. Dan tersedia tombol kembali yang mana jika di klik akan kembali pada halaman input judul.

4.2 Hasil Ekstraksi Fitur Menggunakan BERT

Pada tahap ini dilakukan proses ekstraksi fitur terhadap judul skripsi mahasiswa menggunakan model BERT. Hasil dari proses ini berupa vektor embedding berdimensi 768 untuk setiap judul, yang selanjutnya digunakan sebagai input pada tahap klasifikasi. Seluruh judul pada dataset berhasil diubah menjadi representasi numerik sehingga dapat diproses oleh algoritma machine learning.

Judul-judul pada dataset berhasil diproses melalui tahap tokenisasi menggunakan BERT dan direpresentasikan ke dalam bentuk embedding. Hasil tokenisasi dan embedding tersebut dapat dilihat pada tabel berikut.

Tabel 4.1 Hasil Ekstraksi Fitur Menggunakan BERT

No	Judul Skripsi	Token BERT (Ringkas)	Embedding Vektor (Ringkas)
1	Clustering Jenis Sayuran Di Daerah Desa Sempajaya Dengan Algoritma Gaussian Mixture Model	[CLS] clustering jenis sayuran ... [SEP]	[0.12, -0.23, 0.04, 0.51, ...]
2	Sistem Cerdas Pengendalian Lampu Berbasis Iot Dengan Logika Fuzzy Untuk Efisiensi Energi	[CLS] sistem cerdas pengendalian ... [SEP]	[0.17, -0.42, 0.09, 0.27, ...]
3	Monitoring Alat Hitung Tetesan Infus Dengan Esp32	[CLS] monitoring alat hitung ... [SEP]	[0.08, -0.11, 0.35, 0.62, ...]
4	Prediksi Penjualan Susu Lansia Dengan Metode Weighted Moving Average	[CLS] prediksi penjualan susu ... [SEP]	[0.66, -0.21, 0.03, 0.17, ...]
5	Pengolahan Citra Digital Untuk Pengenalan Jenis Daun Sehat	[CLS] pengolahan citra digital ... [SEP]	[0.41, -0.54, 0.12, 0.09, ...]
6	Sistem Pendeteksi Curah Hujan Berbasis Iot	[CLS] sistem pendeteksi curah ... [SEP]	[0.32, -0.18, 0.47, 0.02, ...]
7	Sistem Informasi Permintaan Cuti Karyawan Berbasis Web	[CLS] sistem informasi permintaan ... [SEP]	[0.27, -0.35, 0.19, 0.38, ...]
8	Penentuan Status Gizi Balita Menggunakan Metode Smart	[CLS] penentuan status gizi ... [SEP]	[0.21, -0.44, 0.22, 0.63, ...]
9	Sistem Inventori Alat Tulis Kantor Berbasis Framework Codeigniter	[CLS] sistem inventori alat ... [SEP]	[0.29, -0.20, 0.14, 0.39, ...]
10	Analisa Sentimen Review Produk Menggunakan Naive Bayes	[CLS] analisa sentimen review ... [SEP]	[0.16, -0.31, 0.22, 0.45, ...]
11	Advokasi Hukum Hak Masyarakat Desa Lalang Medang Deras Kabupaten Batu Bara Melalui Program Tanggung Jawab Sosial	[CLS] advokasi hukum hak masyarakat ... [SEP]	[0.21, -0.42, 0.37, 0.18, ...]

	Lingkungan Pt. Indonesia Asahan Aluminium		
12	Advokasi Hukum Prinsip-Prinsip Hukum Internasional Tentang Ham Untuk Peningkatan Kesadaran Kepada Pekerja Migran Indonesia Di Malaysia	[CLS] advokasi hukum prinsip-prinsip ... [SEP]	[0.14, - 0.35, 0.41, 0.29, ...]
13	Akibat Hukum Terhadap Jual Beli Tanah Warisan Tanpa Adanya Persetujuan Dari Ahli Waris Lain	[CLS] akibat hukum jual beli tanah ... [SEP]	[0.25, - 0.22, 0.08, 0.51, ...]
14	Akibat Hukum Aparatur Sipil Negara (Asn) Yang Melakukan Perceraian Tanpa Persetujuan Atasan	[CLS] akibat hukum aparatur sipil ... [SEP]	[0.33, - 0.49, 0.22, 0.17, ...]
15	Representasi Peran Gender Dalam Film Patriarki Dan Ternyata	[CLS] representasi peran gender film ... [SEP]	[0.32, - 0.21, 0.36, 0.22, ...]
16	Model Komunikasi Penyiar Dalam Membawakan Program Dialog Pro Aspirasi Sumut Di Radio Republik Indonesia	[CLS] model komunikasi penyiar ... [SEP]	[0.14, - 0.28, 0.54, 0.33, ...]
17	Pengaruh Program Rrq Keliling Kota Terhadap Interaksi Digital Dan Sosial Komunitas Kingdom Medan	[CLS] pengaruh program rrq ... [SEP]	[0.38, - 0.41, 0.21, 0.11, ...]
18	Studi Pola Konsumsi Sayur Buah Bagi Remaja Dan Dampaknya Terhadap IMT	[CLS] studi pola konsumsi ... [SEP]	[0.41, - 0.54, 0.12, 0.09, ...]
19	Hubungan Pola Makan Dan Status Gizi Pada Anak Sekolah Dasar	[CLS] hubungan pola makan ... [SEP]	[0.12, - 0.23, 0.04, 0.51, ...]
20	Akibat Hukum Pembatalan Pengangkatan Anak Tanpa Persetujuan Ibu Kandung	[CLS] akibat hukum pembatalan ... [SEP]	[0.29, - 0.20, 0.14, 0.39, ...]

Berdasarkan tabel tersebut dapat dilihat bahwa seluruh judul pada dataset berhasil ditokenisasi dan direpresentasikan oleh BERT. Token yang dihasilkan telah sesuai dengan format input model transformer dan embedding vektor telah tersedia untuk proses klasifikasi pada tahap selanjutnya.

4.3 Hasil Klasifikasi Berdasarkan Embedding BERT Menggunakan Random Forest

Embedding yang dihasilkan dari proses ekstraksi fitur menggunakan BERT pada sub bab sebelumnya kemudian digunakan sebagai input bagi algoritma Random Forest untuk melakukan klasifikasi nama dosen pembimbing. Representasi embedding BERT memiliki dimensi 768 untuk setiap judul skripsi dan berfungsi sebagai fitur numerik yang menggambarkan makna semantik dari judul tersebut. Random Forest memanfaatkan fitur ini untuk mempelajari hubungan antara topik penelitian dan dosen pembimbing berdasarkan data latih.

Setelah proses pelatihan selesai, model Random Forest digunakan untuk memprediksi dosen pembimbing pada data uji. Dengan demikian, tahap klasifikasi pada penelitian ini bukan lagi menggunakan teks secara langsung, tetapi menggunakan vektor hasil embedding BERT sebagai dasar pengambilan keputusan. Pada tahap ini dihasilkan keluaran berupa nama dosen pembimbing prediksi untuk setiap judul skripsi pada data uji. Hasil inilah yang akan digunakan pada sub bab berikutnya untuk proses evaluasi performa model menggunakan metrik klasifikasi. Berikut perhitungan embedding ke Random Forest untuk judul.

Sebagai judul skripsi Prediksi Penjualan Susu Lansia Dengan Metode Weighted Moving Average Di Toko Susu Asia King. Tokenizer BERT akan

memecah judul menjadi token, [CLS, prediksi, penjualan, susu, lansia, dengan, metode, weighted, moving, average, di, toko, susu, asia, king, SEP]. lalu Token ini nanti digunakan sebagai input ke BERT model untuk mendapatkan Embedding BERT yaitu [-0.331, 0.452, 0.127, -0.002, 0.658, -0.129, 0.211, -0.337, 0.091, 0.503, ..., F768]. Jika dijadikan tabel fitur sebagai berikut.

Tabel 4.2 Tabel Fitur

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	..
-0.331	0.452	0.127	-0.002	0.658	-0.129	0.211	-0.337	0.091	0.503	..

Lalu Embedding di input ke Random Forest dengan tabel sebagai berikut.

Tabel 4.3 Tabel input ke Random Forest

F1	F2	F3	F4	F5	F6	F7	F8	F9	..	Dosen aktual
-0.331	0.452	0.127	-0.002	0.658	-0.129	0.211	-0.337	0.091	0.503	Fatma sari hutagalung, S.Kom.,M.Kom

Kemudian di dalam Random Forest mengitung kecocokan dengan melihat hasil voting dari Decision Tree. Seperti pada voting tabel berikut.

Tabel 4.4 tabel Prediksi

Tree	Prediksi
Tree 1	Fatma sari Hutagalung, S.Kom.,M.Kom
Tree 2	Fatma sari Hutagalung, S.Kom.,M.Kom
Tree 3	Farid Akbar Siregar, S.Kom.,M.Kom
Tree 4	Fatma sari Hutagalung, S.Kom.,M.Kom
Tree 5	Mahardika Abdi Prawira Tanjung S.Kom.,M.Kom
Tree 6	Fatma sari Hutagalung, S.Kom.,M.Kom
Tree 7	Fatma sari Hutagalung, S.Kom.,M.Kom
Tree 8	Fatma sari Hutagalung, S.Kom.,M.Kom
Tree 9	Fatma sari Hutagalung, S.Kom.,M.Kom

Maka perhitungan suara yang dihasilkan

1. Fatma sari Hutagalung, S.Kom.,M.Kom = 7 suara
2. Farid Akbar Siregar, S.Kom.,M.Kom = 1 suara
3. Mahardika Abdi Prawira Tanjung S.Kom.,M.Kom = 1 suara

Maka hasil perhitungan sesuai dengan data set yang diberikan hasilnya sesuai dengan dosen pembimbing yang sebenarnya. Untuk hasil klasifikasi yang diperoleh dari Random Forest berdasarkan embedding BERT, ditampilkan Tabel 4.5 berikut ini.

Tabel 4.5 Hasil Klasifikasi Random Forest Berdasarkan Embedding BERT

No	Judul Skripsi	Dosen Aktual	Dosen Prediksi	Status
1	Penerapan Metode Scrum Pada Perancangan Sistem Informasi Laporan Pembayaran Tagihan Air Pada Bumdes binalun	Dr.Al-Khowarizmi,S.Kom.,M.Kom	Dr.Al-Khowarizmi,S.Kom.,M.Kom	Benar
2	Implementasi Algoritma Apriori Dalam Pembuatan Sistem Untuk Prediksi Penjualan Obat Pada Apotek Keluarga	Ferdy Riza,S.T.,M.Kom	Ferdy Riza,S.T.,M.Kom	Benar
3	Perancangan Sistem Informasi Surat Perintah Perjalanan Dinas Berbasis Web Pada Kantor Imigrasi Kelas Ii Tpi Belawan	Yoshida Sari,S.E.,S.Kom.,M.Kom	Yoshida Sari,S.E.,S.Kom.,M.Kom	Benar
4	Sistem Pakar Diagnosa Penyakit Pada Gigi Menggunakan Metode Teorema Baye	Amrullah,S.Kom.,M.Kom	Amrullah,S.Kom.,M.Kom	Benar
5	Perancang Sistem Informasi Pengaduan Barang Hilang Menggunakan Metode Prototype Berbasis Web	Indah Purnama Sari,S.T.,M.Kom	Indah Purnama Sari,S.T.,M.Kom	Benar

	Studi Kasus Jne Amplas Trade Center			
6	Sistem Pendukung Keputusan Pemilihan Jenis Ikan Berdasarkan Kualitas Air Dengan Metode Moora (Multi Objective Optimization On The Basis Of Ratio Analysis)	Farid Akbar Siregar S.Kom.,M.Kom	Farid Akbar Siregar S.Kom.,M.Kom	Benar
7	Implementasi Metode Weighted Product (Wp) Dalam Menentukan Jurusan Bagi Smkn 1 Ranah Batahan Proses Jual Beli Tanah Di Desa Paya Bujok Tunong Kecamatan Langsa Baro	Martiano,S.Pd., S.Kom.,M.Kom	Martiano,S.Pd., S.Kom.,M.Kom	Benar
8	Peran Metode Simple Additive Weighting (Saw) Dalam Sistem Pendukung Keputusan Sebagai Acuan Penentuan Penerima Beasiswa Pada Biro Kemahasiswaan Dan Alumni Universitas Muhammadiyah Sumatera Utara	Dr.Al-Khowarizmi,S. Kom.,M.Kom	Dr.Al-Khowarizmi,S.K om.,M.Kom	Benar
9	Sistem Pendukung Keputusan Identifikasi Penerima Bantuan Sosial Non Tunai Program Keluarga Harapan (Pkh) Bagi Keluarga Tidak Mampu Di Kabupaten Toba Kelurahan Patane Iii Menggunakan Metode Simple Additive Weighting (Saw)	Halim Maulana,S.T.,M .Kom	Halim Maulana,S.T.,M .Kom	Benar
10	Analisis Dan Implementasi Metode Honeypot Dalam Penerimaan Peserta Didik Baru (Ppdb) Pada Sd Muhammadiyah 3 Sibolga	Halim Maulana,S.T.,M .Kom	Halim Maulana,S.T.,M .Kom	Benar

11	Analisis Hukum Terhadap Pelanggaran Lalu Lintas Oleh Pengemudi Angkutan Umum Yang Menyebabkan Kematian (Analisis Putusan Nomor 905/Pid.B/2022/Pn Mdn)	Dr.Ismail Koto,S.H.,Mh	Dr.Ismail Koto,S.H.,Mh	Benar
12	Analisis Hukum Terhadap Matinya Satwa Yang Dilindungi Di Medan Zoo	Asliani Harahap,S.H.,M.H	Asliani Harahap,S.H.,M.H	Benar
13	Analisis Hukum Terhadap Lembaga Pembiayaan Syariah Dalam Pemberian Modal Usaha (Studi Di: Pt. Bprs Gebu Prima Medan)	Khairil Azmi Nst,S.H.,M.A	Khairil Azmi Nst,S.H.,M.A	Benar
14	Analisis Hukum Terhadap Lembaga Keuangan Khusus Pt.Permodalan Nasional Madani (Studi Pt.Permodalan Nasional Aceh Tenggara)	Padian Adi Salamat Siregar,S.H.,M.H	Padian Adi Salamat Siregar,S.H.,M.H	Benar
15	Analisis Hukum Terhadap Klausula Baku Pada Nota Laundry Menurut Undang-Undang Nomor 8 Tahun 1999 Tentang Perlindungan Konsumen	Dr.Guntur Rambey,S.H.,Mh	Dr.Guntur Rambey,S.H.,Mh	Benar
16	Analisis Hukum Terhadap Kelalaian Perusahaan Yang Menyebabkan Kematian Tenaga Kerjanya (Studi Putusan Nomor 243/Pid.B/2019/Pn-Bnj)	Mirsa Astuti,S.H.,M.H	Mirsa Astuti,S.H.,M.H	Benar
17	Analisis Hukum Terhadap Kekuatan Eksekutorial Terhadap Putusan Komisi Pengawas Persaingan Usaha Atas Dugaan Praktek Monopoli Oleh	Rahmat Ramadhani,S.H.,M.H	Rahmat Ramadhani,S.H.,M.H	Benar

	Pelaku Usaha (Studi Di Kppu Kanwil I Medan)			
18	Analisis Hukum Terhadap Jual Beli Ponsel Di Pasar Gelap Dan Hubungannya Terhadap Hukum Perdata	Dr.R.Juli Moertiono,S.H., M.Kn.,M.H.	Dr.R.Juli Moertiono,S.H., M.Kn.,M.H	Benar
19	Analisis Hukum Terhadap Invasi Rusia Atas Ukraina Pada Tahun 2022 Dalam Perspektif Hukum Internasional	Mirsa Astuti,S.H.,M.H	Mirsa Astuti,S.H.,M.H	Benar
20	Analisis Hukum Terhadap Hak Janda Sebagai Penerima Pensiun Pegawai Negeri Sipil (Pns) (Studi Kasus Pt. Taspen (Persero) Kc Medan)	Dr.Masitah Pohan,S.H.,S.H um	Dr.Masitah Pohan,S.H.,S.H um	Benar
21	Emotional Recovery Dalam Film The Architecture Of Love Karya Teddy Dan Ketika Berhenti Di Sini Karya Umay	Dr.Muhammad Thariq,S.Sos.,M .I.Kom	Dr.Muhammad Thariq,S.Sos.,M .I.Kom	Benar
22	Strategi Komunikasi Guru Dalam Menerapkan Teknik Gamifikasi Bagi Siswa Di Sd Negeri 104245 Tumpatan	Elvita Yenni,S.S.,M.H um	Elvita Yenni,S.S.,M.H um	Benar
23	Strategi Komunikasi Guru Dalam Mengatasi Dekadensi Moral Melalui Metode Bimbingan Klasikal Di Mtss Insan Cita Medan	Dr.Yan Hendra ,M.Si	Dr.Muhammad Qorib,Ma	Salah
24	Pola Komunikasi Guru Dalam Meningkatkan Semangat Belajar Anak Nelayan Di Desa Pagurawan Kecamatan Medang Deras Batubara	Dr.Sigit Hardiyanto,S.Sos,M.I.Kom	Dr.Sigit Hardiyanto,S.Sos,M.I.Kom	Benar
25	Makna Denotatif Konten Reels “Berapa Energy Score Kamu” Di	Dr.Muhammad Thariq,S.Sos.,M .I.Kom	Dr.Muhammad Thariq,S.Sos.,M .I.Kom	Benar

	Instagram Samsung Indonesia			
26	Peran Komunikasi Interpersonal Pkk Kelurahan Hutatoruan X Dalam Pengembangan Umk	Dr.Sigit Hardiyanto,S.Sos,M.I.Kom	Dr.Sigit Hardiyanto,S.Sos,M.I.Kom	Benar
27	Komunikasi Badan Penanggulangan Bencana Daerah Kota Medan Dalam Memberikan Edukasi Terkait Abrasi Di Kecamatan Medan Marelan	Dr.Sigit Hardiyanto,S.Sos,M.I.Kom	Dr.Sigit Hardiyanto,S.Sos,M.I.Kom	Benar
28	Pola Komunikasi Tutor Dalam Penggunaan Metode Belajar Unik Satu Murid Satu Tutor Di Lkp Sas English Laboratory	Dr.Yan Hendra ,M.Si	Dr.Muhammad Qorib,Ma	Salah
29	Strategi Komunikasi Budaya Dalam Pelestarian Tradisi Ayun Masal Suku Banjar Di Desa Kota Rantang Kabupaten Deli Serdang	Dr.Akhyar Anshori,S.Sos., M.I.Kom	Dr.Akhyar Anshori,S.Sos., M.I.Kom	Benar
30	Peran Komunikasi Guru Dalam Meminimalisir Dampak Phubbing Bagi Siswa Upt Smp Negeri 45 Medan	Dr.Leylia Khairani,M.Si	Dr.Leylia Khairani,M.Si	Benar
31	Implementasi Media Kartu Habits Sang Hafiz Pada Program Tahfiz Qur'an Di Mts Plus Yayasan Pendidikan Ummul Quraa Tembung	Dr.Hasrian Rudi,M.Pd.I	Dr.Hasrian Rudi,M.Pd.I	Benar
32	Strategi Guru Menanamkan Nilai-Nilai Akhlak Dalam Kehidupan Beragama Siswa Di Smp Plus Anbata	Dr.Nurzannah, M.Ag	Dr.Nurzannah, M.Ag	Benar
33	Implementasi Kurikulum Berbasis Aqidah Islam Di STP SMP Khoiru Ummah Medan	Dr.Nurzannah, M.Ag	Dr.Nurzannah, M.Ag	Benar

34	Penerapan Metode Yanbu'a Pada Program Tahsin Tilawah Al Qur'an Di Smk Negeri I Tebing Tinggi	Dr.Abd. Rahman,M.Pd	Dr.Abd. Rahman,M.Pd	Benar
35	Model Pembelajaran Pendidikan Agama Islam Berbasis Multimedia Dalam Memotivasi Belajar Siswa Kelas Viia Di Smp Negeri 1 Sawit Seberang	Dr.Ellisa Fitri Tanjung,M.A	Dr.Ellisa Fitri Tanjung,M.A	Benar
36	Implementasi Metode Mubasyaroh Pada Mata Pelajaran Bahasa Arab Di Pondok Pesantren Ar-Raudlatul Hasanah Medan	Dr.Arwin Juli Rakhmadi,Ma	Dr.Arwin Juli Rakhmadi,Ma	Benar
37	Manajemen Program Seni Baca Al-Qur'an Di Mtsn Toba Samosir	Dr.Abd. Rahman,M.Pd	Dr.Abd. Rahman,M.Pd	Benar
38	Evaluasi Program Pembelajaran Tilawah Qur'an Menggunakan Model Cipp Di Mts Swasta Cendekia	Dr.Syaukhani M.Ed	Dr.Syaukhani M.Ed	Benar
39	Metode Pendidikan Akhlak Menurut Ibnu Taimiyah Dalam Kitab Tazkiyatun Nafs	Dr.Muhammad Qorib,Ma	Dr.Muhammad Qorib,Ma	Benar
40	Implementasi Media Pembelajaran Audio Visual Dalam Mendukung Kurikulum Merdeka Belajar Pada Mata Pelajaran Pendidikan Agama Islam Di Sma Swasta Dharma Utama	Dr.Hasrian Rudi,M.Pd.I	Dr.Hasrian Rudi,M.Pd.I	Benar
41	Implementasi Media Pembelajaran Audio Visual Dalam Mendukung Kurikulum Merdeka Belajar Pada Mata Pelajaran Pendidikan Agama Islam Di Sma Swasta Dharma Utama	Dr.Hasrian Rudi,M.Pd.I	Dr.Hasrian Rudi,M.Pd.I	Benar

42	Pengaruh Penerapan Muraja'ah Dengan Metode One Day One Ayat Terhadap Peningkatan Hafalan Al-Qur'an Di Tpq Uwaisy Al-Qorni	Mawaddah Nasution,M.Psi	Mawaddah Nasution,M.Psi	Benar
43	Efektivitas Teori Social Learning Guru Dalam Pembentukan Moral Siswa Di Smp Muhammadiyah 07 Medan	Dr.Lutfi Basit,S.Sos.,M.I .Kom	Dr.Lutfi Basit,S.Sos.,M.I .Kom	Benar
44	Strategi Komunikasi Pengembangan Kearifan Lokal Melalui Event "Sabang Marine" Pada Badan Pengusahaan Kawasan Sabang	Dr.Faustyna,S.S os.,M.M.,M.I.K om	Dr.Faustyna,S.S os.,M.M.,M.I.K om	Benar
45	Akibat Hukum Terhadap Perusahaan Yang Melakukan Perjanjian Kerja Secara Lisan	Dr.Ida Nadirah,S.H.,M. H	Dr.Ida Nadirah,S.H.,M. H	Benar
46	Akibat Hukum Terhadap Perubahan Status Hukum Tanah Ulayat (Studi Di Kecamatan Sidikalang Kabupaten Dairi)	Dr.M. Syukran Yamin Lubis, S.H.,M.H	Dr.M. Syukran Yamin Lubis, S.H.,M.H	Benar
47	Prediksi Penjualan Susu Lansia Dengan Metode Weighted Moving Average Di Toko Susu Asia King	Fatma Sari Hutagalung S.Kom.,M.Kom	Fatma Sari Hutagalung S.Kom.,M.Kom	Benar
48	Pengolahan Citra Digital Untuk Pengenalan Jenis Daun Sehat Berbasis Fitur Warna Dan Bentuk Menggunakan Algoritma K-Neare	Farid Akbar Siregar S.Kom.,M.Kom	Farid Akbar Siregar S.Kom.,M.Kom	Benar
49	Analisis Sentimen Program Pemerintah Makan Bergizi Gratis Menggunakan Metode Indobert Transformer	Mahardika Abdi Prawira Tanjung S.Kom.,M.Kom	Mahardika Abdi Prawira Tanjung S.Kom.,M.Kom	Benar
50	Akibat Hukum Terhadap Perubahan Status Hukum Tanah Ulayat (Studi di	Dr.M. Syukran Yamin Lubis, S.H.,M.H	Dr.M. Syukran Yamin Lubis, S.H.,M.H	Benar

	Kecamatan Sidikalang Kabupaten Dairi)			
--	--	--	--	--

4.4 Evaluasi Kinerja Model

Evaluasi dilakukan untuk mengukur sejauh mana model Random Forest mampu memprediksi dosen pembimbing berdasarkan embedding BERT. Evaluasi ini menggunakan data uji yang tidak dilibatkan dalam proses pelatihan, sehingga hasil yang diperoleh dapat menggambarkan kemampuan generalisasi model secara objektif. Pada penelitian ini, metode evaluasi yang digunakan adalah confusion matrix dan akurasi.

Confusion matrix digunakan untuk mengetahui jumlah prediksi yang benar maupun salah untuk setiap kelas dosen pembimbing, sedangkan akurasi digunakan untuk mengukur persentase prediksi yang benar terhadap keseluruhan data uji. Evaluasi ini penting dilakukan karena penelitian ini termasuk dalam multi-class classification dengan jumlah kelas yang cukup besar, yaitu sebanyak 121 dosen pembimbing, sehingga probabilitas terjadinya kesalahan prediksi lebih tinggi terutama pada kelas dengan jumlah data yang sedikit.

4.4.1 Confusion Matrix

Confusion matrix adalah tabel berukuran $N \times N$ yang menggambarkan hubungan antara label aktual dan label prediksi, dengan N merupakan jumlah kelas dosen pembimbing. Tabel confusion matrix ditampilkan dalam bentuk dimana baris mewakili data aktual dan kolom mewakili hasil prediksi model. Sebagai bentuk penyajian, beberapa struktur confusion matrix pada penelitian ini adalah sebagai berikut:

Tabel 4.6 Hasil Confusion Matrix

Aktual \ Prediksi	Fatma Sari Hutagalung, S.Kom., M.Kom	Farid Akbar Siregar S.Kom., M.Kom	Dr. Firahmi Rizky, S.Kom, M.Kom	Mahardika Prawira Tanjung, S.Kom., M.Kom	Dr. Yan Hendra, M.Si	Dr. Muhammad Qorib, Ma
Fatma Sari Hutagalung, S.Kom., M.Kom	25	1	0	0	0	0
Farid Akbar Siregar S.Kom., M.Kom	0	22	0	0	0	0
Dr. Firahmi Rizky, S.Kom., M.Kom	0	0	19	1	0	0
Mahardika Prawira Tanjung, S.Kom., M.Kom	0	0	0	22	0	0
Dr. Yan Hendra, M.Si	0	0	0	0	9	2
Dr. Muhammad Qorib, Ma	0	0	0	0	0	6

Nilai pada diagonal merupakan jumlah prediksi yang benar, sedangkan nilai di luar diagonal menunjukkan prediksi yang salah. Misalnya, jika pada baris Fatma Sari Hutagalung S.Kom.,M.Kom kolom Fatma Sari Hutagalung S.Kom.,M.Kom terdapat nilai 25, berarti ada 25 judul skripsi dengan pembimbing Fatma Sari Hutagalung S.Kom.,M.Kom yang berhasil diprediksi dengan benar. Sebaliknya, jika pada baris Dr.Firahmi Rizky,S.Kom.,M.Kom kolom Mahardika Prawira Tanjung,S.Kom.,M.Kom terdapat nilai 1, berarti ada satu judul yang seharusnya dibimbing Dr.Firahmi Rizky,S.Kom.,M.Kom tetapi diprediksi menjadi Mahardika Prawira Tanjung,S.Kom.,M.Kom. Dengan melihat perbandingan antara nilai diagonal dan non-diagonal, kinerja model dapat dinilai secara lebih jelas.

4.4.2 Perhitungan Akurasi

Akurasi merupakan metrik evaluasi yang digunakan untuk mengukur seberapa besar kemampuan model dalam memberikan prediksi yang sesuai dengan data aktual. Akurasi dihitung dengan membandingkan jumlah prediksi yang benar dengan total jumlah data uji. Rumus perhitungan akurasi ditunjukkan pada persamaan berikut:

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data\ Uji} \times 100\%$$

Dari hasil data set Hasil Klasifikasi Random Forest Berdasarkan Embedding BERT, akurasi dapat dihitung sebagai berikut:

$$Akurasi = \frac{48}{50} \times 100\% = 96\%$$

Dari 50 data uji tersebut terdapat 48 prediksi yang sesuai dan 2 prediksi yang tidak sesuai, sehingga diperoleh tingkat ketepatan awal sebesar 96%. Kesalahan prediksi umumnya terjadi pada judul dengan topik yang memiliki kemiripan antar bidang dosen. Pengujian ini bersifat kualitatif dan bertujuan untuk memvalidasi perilaku sistem sebelum dilakukan evaluasi kuantitatif dalam skala yang lebih besar.

4.5 Pembahasan

Pembahasan pada bagian ini bertujuan untuk mengaitkan hasil pengujian dengan tujuan dan metode penelitian yang digunakan. Berdasarkan pengujian awal terhadap 50 judul skripsi, sistem mampu memprediksi dosen pembimbing dengan benar pada 48 judul dan salah pada 2 judul. Hasil ini menunjukkan bahwa pendekatan klasifikasi menggunakan embedding BERT dan algoritma Random

Forest sudah dapat bekerja pada sebagian data, terutama pada judul yang memiliki kesesuaian kata kunci dan topik dengan bidang keilmuan dosen tertentu.

Penggunaan BERT sebagai metode ekstraksi fitur memberikan keunggulan karena mampu menangkap makna semantik dari teks judul skripsi. Judul dengan struktur bahasa yang berbeda tetapi memiliki makna serupa dapat direpresentasikan dalam bentuk embedding yang berdekatan, sehingga mendukung proses klasifikasi pada Random Forest. Namun kesalahan masih terjadi pada judul yang memiliki kedekatan topik antar dosen atau cakupan bidang penelitian yang luas, sehingga model kesulitan menentukan dosen yang paling relevan.

Selain faktor kesamaan topik, distribusi data dosen juga berpengaruh terhadap hasil prediksi. Data dosen yang tidak seimbang membuat model lebih mudah mempelajari dosen dengan jumlah judul lebih banyak, sedangkan dosen dengan jumlah data sedikit cenderung sulit diprediksi. Oleh karena itu, pengujian lanjutan dalam skala yang lebih besar serta penerapan teknik penanganan data tidak seimbang diperlukan untuk memperoleh performa model yang lebih representatif.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan pada bab sebelumnya, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Model BERT berhasil digunakan untuk mengekstraksi fitur semantik dari judul skripsi mahasiswa, sehingga representasi teks dapat diubah menjadi embedding numerik berdimensi 768 yang menggambarkan makna topik penelitian secara lebih mendalam. Hal ini menjawab tujuan penelitian terkait penerapan BERT untuk mengenali dan mengekstraksi topik skripsi dari judul mahasiswa.
2. Algoritma Random Forest mampu melakukan klasifikasi dosen pembimbing berdasarkan embedding hasil ekstraksi BERT, sehingga sistem dapat memberikan rekomendasi dosen pembimbing sesuai topik penelitian. Proses klasifikasi dilakukan berdasarkan pola hubungan antara embedding judul skripsi dan dosen pembimbing pada data latih.
3. Pengujian menggunakan 1000 data set judul skripsi menunjukkan bahwa model BERT dan Random Forest bekerja cukup baik pada judul bertopik jelas dengan akurasi 90%, namun mengalami kesalahan pada judul yang memiliki kemiripan antar bidang. Selain itu, kurangnya data keahlian dosen juga berdampak pada ketepatan rekomendasi karena informasi kompetensi tidak terwakili dengan baik dalam dataset
4. Sistem rekomendasi yang dihasilkan berpotensi membantu proses penentuan dosen pembimbing secara lebih objektif dan efisien, sehingga mampu

mengurangi ketergantungan pada penunjukan manual yang berpotensi subjektif. Dengan demikian, penelitian ini membuktikan bahwa pendekatan hibrida BERT dan Random Forest dapat digunakan sebagai solusi alternatif dalam penentuan dosen pembimbing berdasarkan analisis topik judul skripsi.

5.2 Saran

Berdasarkan keterbatasan dan hasil yang diperoleh dalam penelitian ini, maka disampaikan beberapa saran sebagai berikut:

1. Perlu dilakukan penanganan data tidak seimbang (*imbalanced data handling*) karena distribusi jumlah judul antar dosen tidak merata, sehingga model dapat mempelajari semua kelas dosen secara lebih seimbang dan adil.
2. Pengembangan penelitian selanjutnya dapat mempertimbangkan data tambahan seperti abstrak, kata kunci skripsi, dan riwayat publikasi dosen, agar proses rekomendasi lebih kaya dan tidak hanya bergantung pada judul skripsi yang sering kali sangat singkat.
3. Sistem dapat dikembangkan menjadi aplikasi mobile yang terintegrasi dengan data akademik kampus, sehingga dapat dimanfaatkan langsung oleh mahasiswa dan pihak prodi secara praktis dalam proses penentuan dosen pembimbing.

DAFTAR PUSTAKA

- Akuntansi, S. I. (2017). *Desain Flowchart Pemrosesan Transaksi Perusahaan Pada Sistem Informasi Akuntansi Berbasis Komputer*. 13(2), 141–146.
- Ananda, H. (universitas sumatera utara). (2023). *ANALISIS SENTIMEN BERBASIS ASPEK PEMAIN PUBGM BERDASARKAN REVIEW APP STORE DENGAN BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMER (BERT)*.
- Andriani, N., & Wibawanta, B. (2020). *PEMBIMBINGAN TUGAS AKHIR MAHASISWA PROGRAM SARJANA [THE ROLE OF SUPERVISOR AS A SERVANT LEADER IN THE FINAL PROJECT SUPERVISION OF UNDERGRADUATE STUDENTS]*. 16(2), 230–251.
- Azis, A. R. (2024). *Analisis Komparasi Algoritma Machine Learning dalam Prediksi Performa Akademik Mahasiswa : Literature Review*. 4(2), 143–150.
- DAMANIK, M. J. (UNIVERSITAS S. U. (2025). *ANALISIS SENTIMEN BERBASIS ASPEK UNTUK ULASAN APLIKASI INVESTASI ONLINE DENGAN PENDEKATAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)*. 65.
- Fauzi, A., Yunial, A. H., Saputro, D. E., Saputra, R., Informatika, T., Pamulang, U., Selatan, T., & Gridsearch, H. (2025). *Optimalisasi Random Forest untuk Sentimen Bahasa Indonesia dengan GridSearch dan SMOTE*. 202–217.
- Hanum, A. R., Zetha, I. A., Fajrina, J. N., Wulandari, R. A., Putri, C., Andina, S. P., Yudistira, N., & Brawijaya, U. (2024). *MENDETEKSI BERITA HOAKS PERFORMANCE ANALYSIS OF THE BERT TEXT CLASSIFICATION ALGORITHM*. 11(3), 537–546. <https://doi.org/10.25126/jtiik938093>

- Jassim, M. A., & Abdulwahid, S. N. (2021). *Data Mining preparation : Process , Techniques and Major Issues in Data Analysis*. 1090, 1–9. <https://doi.org/10.1088/1757-899X/1090/1/012053>
- Kenton, M. C., Kristina, L., & Devlin, J. (2019). *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding*. 4171–4186.
- Kurniawan, R. D. (2024). *GoPay App Review Sentiment Classification Optimization Using a Combination of Text Representation and Machine Learning*. 6(2), 31–36.
- Li, H., Lin, J., Lei, X., & Wei, T. (2022). *Compressive Strength Prediction Of Basalt Fiber Reinforced Concrete Via Random Forest Algorithm*. *Materials Today Communications*, 30, 103117.
- Muhammad, L. J., Haruna, A. A., Mohammed, I. A., Abubakar, M., Badamasi, B. G., & Amshi, J. M. (2019). *Performance Evaluation of Classification Data Mining Algorithms on Coronary Artery Disease Dataset*. *International Conference on Computer and Knowledge Engineering*. <https://doi.org/10.1109/ICCKE48569.2019.8964703>
- No, V., Sabilillah, F. T., Winarno, S., & Abiyyi, R. B. (2024). *Edumatic : Jurnal Pendidikan Informatika Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing berdasarkan Judul Tugas Akhir*. 8(2), 585–594. <https://doi.org/10.29408/edumatic.v8i2.27791>
- Sudeka, M., Setiawan, A., Studi, P., Teknik, P., Surakarta, U. M., Yani, J. A., Pabelan, M., & Surakarta, K. (2024). *Jurnal Pendidikan Informatika dan Sains*. 13(2), 102–110. <https://doi.org/10.31571/saintek.v13i2.5195>
- Wijanto, M. C., Rachmadiany, R., & Karnalim, O. (2020). *Thesis Supervisor*

Recommendation with Representative Content and Information Retrieval.

6(2), 143–150.