



SKRIPSI

**PENERAPAN CONTENT-BASED FILTERING DAN
NAÏVE BAYES CLASSIFIER PADA REKOMENDASI
DOSEN PEMBIMBING TUGAS AKHIR**

BINTANG JAGAD SYAH SANA
NPM 21081010005

DOSEN PEMBIMBING

Eva Yulia Puspaningrum, S.Kom., M.Kom
Muhammad Muharrom Al Haromainy, S.Kom, M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

PENERAPAN CONTENT-BASED FILTERING DAN NAÏVE BAYES CLASSIFIER PADA REKOMENDASI DOSEN PEMBIMBING TUGAS AKHIR

BINTANG JAGAD SYAH SANA

NPM 21081010005

DOSEN PEMBIMBING

Eva Yulia Puspaningrum, S.Kom., M.Kom

Muhammad Muharrom Al Haromainy, S.Kom, M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

LEMBAR PENGESAHAN

PENERAPAN CONTENT-BASED FILTERING DAN NAÏVE BAYES CLASSIFIER PADA REKOMENDASI DOSEN PEMBIMBING TUGAS AKHIR

Oleh :
BINTANG JAGAD SYAH SANA
NPM.21081010005

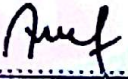
Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 12 September 2025

Menyetujui

Eva Yulia Puspaningrum, S.Kom., M.Kom
NIP. 19890705 202121 2 002


..... (Pembimbing I)

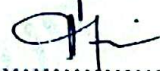
M. Muharrom Al Haromainy, S.Kom., M.Kom
NIP. 19950601 202203 1 006


..... (Pembimbing II)

Yisti Vita Via, S.ST., M.Kom.
NIP. 19860425 202121 2 001


..... (Ketua Penguji)

Afina Lina Nurlaili, S.Kom., M.Kom.
NIP. 1993121 3202203 2 010


..... (Anggota Penguji)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**PENERAPAN CONTENT-BASED FILTERING DAN NAÏVE BAYES
CLASSIFIER PADA REKOMENDASI DOSEN PEMBIMBING TUGAS
AKHIR**

Oleh :
BINTANG JAGAD SYAH SANA
NPM. 21081010005

Telah disetujui untuk mengikuti Ujian Skripsi



Menyetujui,
**Koordinator Program Studi Informatika
Fakultas Ilmu Komputer**

Fetty Tri Anggraeny, S.Kom., M.Kom.
NIP 19820211 2021212 005

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan di bawah ini:

Nama Mahasiswa : Bintang Jagad Syah Sana
Tempat Tanggal Lahir : Kediri, 29 Agustus 2002
NPM : 21081010005
Bidang/Konsentrasi : Informatika/SCR
Alamat : Ds. Kedungsari Kec.Tarokan Kab.Kediri

Dengan ini menyatakan dengan sebenar-benarnya bahwa skripsi berjudul **“PENERAPAN CONTENT-BASED FILTERING DAN NAÏVE BAYES CLASSIFIER PADA REKOMENDASI DOSEN PEMBIMBING TUGAS AKHIR”** adalah benar-benar karya asli saya sendiri dan tidak pernah diajukan untuk memperoleh gelar akademik apa pun di perguruan tinggi mana pun.

Segala kutipan atau pendapat orang lain yang tercantum dalam skripsi ini telah disebutkan sumbernya sesuai ketentuan penulisan karya ilmiah. Apabila di kemudian hari terbukti pernyataan ini tidak benar, saya bersedia menanggung segala akibat hukum dan sanksi yang berlaku, termasuk pencabutan gelar yang telah diperoleh. Demikian surat pernyataan ini saya buat dengan penuh tanggung jawab dan kesadaran.

Surabaya, 11 September 2025

Yang Membuat Pernyataan,



Bintang Jagad Syah Sana

NPM. 21081010005

ABSTRAK

Nama mahasiswa / NPM : Bintang Jagad Syah Sana
Judul Skripsi : Penerapan Content-Based Filtering dan Naïve Bayes Classifier pada Rekomendasi Dosen Pembimbing Tugas Akhir
Dosen Pembimbing : 1. Eva Yulia Puspaningrum, S.Kom., M.Kom
2. M. Muharrom Al Haromainy, S.Kom., M.Kom

Pemilihan dosen pembimbing tugas akhir merupakan faktor krusial yang memengaruhi keberhasilan mahasiswa dalam menyelesaikan skripsi. Keterbatasan informasi mengenai kesesuaian bidang keahlian dosen sering menimbulkan ketidaktepatan pemilihan pembimbing. Penelitian ini mengusulkan sistem rekomendasi dosen pembimbing berbasis text mining dengan mengintegrasikan Content-Based Filtering (CBF) dan Naïve Bayes Classifier (NBC). Korpus data mencakup publikasi ilmiah dosen sebagai proksi bidang keahlian serta kumpulan judul skripsi mahasiswa periode 2022–2024 sebagai data uji. Tahap preprocessing meliputi translasi, case folding, tokenisasi, stopword removal, dan stemming. CBF menghitung kemiripan judul–publikasi menggunakan TF-IDF dan cosine similarity, sementara NBC memodelkan probabilitas kecocokan topik dengan Laplace smoothing. Skor kedua metode digabung melalui weighted sum untuk menghasilkan peringkat rekomendasi. Evaluasi menunjukkan pendekatan gabungan CBF dan NBC memberikan kinerja lebih baik dibanding metode tunggal, dengan tren peningkatan akurasi antar-tahun. Secara khusus, pada data validasi tahun 2024, sistem mencapai akurasi 86% pada skema pembobotan optimal, menegaskan efektivitas kombinasi fitur kemiripan konten dan probabilitas topik dalam memetakan kecocokan dosen–mahasiswa. Temuan ini berimplikasi pada proses pemilihan pembimbing yang lebih objektif, transparan, dan terdistribusi, sekaligus menawarkan kerangka yang dapat direplikasi untuk program studi lain dengan karakteristik serupa.

Kata Kunci : Rekomendasi, *Content-Based Filtering*, *Naïve Bayes Classifier*, TF-IDF, *Cosine Similarity*, *Text Mining*

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Bintang Jagad Syah Sana
Thesis Title : Implementation of Content-Based Filtering and
Naïve Bayes Classifier in Final Project Supervisor
Recommendation
Advisor : 1. Eva Yulia Puspaningrum, S.Kom., M.Kom
2. M. Muharrom Al Haromainy, S.Kom., M.Kom

The selection of thesis supervisors is a crucial factor that significantly influences students' success in completing their undergraduate theses. Limited access to information regarding the alignment of lecturers' expertise often leads to inaccurate supervisor assignments. This study proposes a supervisor recommendation system based on text mining by integrating Content-Based Filtering (CBF) and the Naïve Bayes Classifier (NBC). The dataset consists of lecturers' scientific publications as proxies for their expertise and a collection of student thesis titles from 2022–2024 as test data. The preprocessing stage includes translation, case folding, tokenization, stopwords removal, and stemming. CBF measures the similarity between thesis titles and publications using TF-IDF and cosine similarity, while NBC models the probability of topic relevance through Laplace smoothing. The scores from both methods are combined using a weighted sum approach to generate the final ranking of recommended supervisors. The evaluation results indicate that the hybrid CBF+NBC approach outperforms each method individually, showing a consistent improvement in accuracy across the years. Specifically, for the 2024 validation dataset, the system achieved 86% accuracy under the optimal weighting scheme, highlighting the effectiveness of combining content similarity and probabilistic modeling in mapping the compatibility between lecturers and students. This study provides a more objective, transparent, and balanced framework for supervisor selection, and it can be replicated in other academic programs with similar characteristics.

Keywords : Recommendation, Content-Based Filtering, Naïve Bayes Classifier, TF-IDF, Cosine Similarity, Text Mining.

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Penerapan Content-Based Filtering dan Naïve Bayes Classifier pada Rekomendasi Dosen Pembimbing Tugas Akhir”** dapat terselesaikan dengan baik.

Penulis menyadari bahwa penyusunan skripsi ini tidak lepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT, karena telah memberikan rezeki berupa kesehatan, pengetahuan, dan keberuntungan kepada penulis.
2. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom dan Bapak M. Muharrom Al Haromainy, S.Kom., M.Kom, selaku pembimbing skripsi yang telah memberikan arahan, bimbingan dan motivasi selama penyusunan skripsi
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom. M.Kom., selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Ibu Yisti Vita Via, S.ST., M.Kom. dan Ibu Afina Lina Nurlaili, S.Kom., M.Kom., selaku dosen penguji yang telah memberikan masukan dan kritik yang membangun dalam rangka penyempurnaan skripsi ini.
6. Kepada orang tua penulis, Bapak Much. Sujud dan Ibu Supiyah, yang senantiasa mendukung penulis dan mendoakan yang terbaik untuk penulis.
7. Kepada saudara dan keluarga besar penulis, karena selalu memastikan progress skripsi berjalan tanpa harus mem-*pressure* penulis.
8. Kepada Frida Putri Purgito karena telah menemani dalam pengerjaan skripsi dan memberikan emotional support terhadap penulis, serta selalu percaya dengan perkembangan penulis setiap harinya.

9. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan kontribusi baik langsung maupun tidak langsung dalam penyelesaian skripsi ini

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 10 September 2025

Penulis

Bintang Jagad Syah Sana

DAFTAR ISI

LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN ORISINALITAS	v
ABSTRAK	vii
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Penelitian Terdahulu	7
2.2 Landasan Teori.....	10
2.2.1 <i>Content Based Filtering</i>	10
2.2.2 <i>Text Mining</i>	12
2.2.3 Naïve Bayes Classifier	13
2.2.4 Model Rekomendasi	14
2.2.5 NLTK.....	15
2.2.6 Panda	15
2.2.7 Numpy.....	16
2.2.8 Sastrawati	16
2.2.9 <i>Google Translate API</i>	16
2.2.10 <i>Weigted Sum Model</i>	17
2.2.11 <i>Bayes Rule Normalization</i>	18
BAB III DESAIN DAN IMPLEMENTASI SISTEM	19

3.1 Alur Penelitian	19
3.2 Studi Literatur	19
3.3 Pengumpulan Data	20
3.3.1 Data Jurnal Dosen	20
3.3.2 Data Skripsi Mahasiswa	22
3.4 <i>Preprocessing Data</i>	23
3.4.1 <i>Translate</i>	24
3.4.2 <i>Punctional Data</i>	25
3.4.3 <i>Case Folding</i>	26
3.4.4 <i>Tokenization</i>	27
3.4.5 <i>Stopword Removal</i>	28
3.4.6 <i>Stemming</i>	29
3.5 <i>Content Based Filtering</i>	30
3.5.1 TF-IDF	30
3.5.2 <i>Cosine Similarity</i>	36
3.6 <i>Naive Bayes Classifier</i>	38
3.7 Tahap Rekomendasi	43
3.8 Pengujian	46
3.9 Evaluasi	47
BAB IV	49
4.1 Implementasi Metode	49
4.1.1 <i>Import Library</i>	49
4.1.2 <i>Preprocessing</i>	51
4.1.3 Normalisasi Nama Dosen	61
4.1.4 Pembobotan Penulis Pertama	62
4.1.5 <i>Content Bases Filtering</i>	63
4.1.6 <i>Naïve Bayes Classifier</i>	72
4.1.7 Pembobotan Metode	77
4.2 Pengujian Dan Evaluasi	80
4.2.1 Pengujian dan Evaluasi Skripsi 2024	81
4.2.2 Pengujian dan Evaluasi Skripsi 2023	83
4.2.3 Pengujian dan Evaluasi Skripsi 2022	85

4.2.4 Pengujian dan Evaluasi Jurnal Dosen	87
4.3 Hasil Uji Metode	92
4.3.1 Hasil Uji untuk Judul Skripsi 1	92
4.3.2 Hasil Uji untuk Judul S2	100
BAB V KESIMPULAN DAN SARAN	109
5.1 Kesimpulan	109
5.2 Saran.....	110
DAFTAR PUSTAKA	111
LAMPIRAN.....	114

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 3. 1 <i>Flowchart</i> Alur Penelitian	19
Gambar 3. 2 <i>Flowchart Preprocessing</i>	24
Gambar 3. 3 <i>Flowchart Content Based Filtering</i>	30
Gambar 3. 4 <i>Flowchart Naive Bayes Classifier</i>	38
Gambar 4. 1 Display Data Jurnal Dosen	52
Gambar 4. 2 Hasil <i>Translate</i>	53
Gambar 4. 3 Hasil <i>Lower Case</i>	55
Gambar 4. 4 Hasil <i>Punctional Data</i>	56
Gambar 4. 5 Hasil Tokenization	57
Gambar 4. 6 Hasil Stopwords	58
Gambar 4. 7 Hasil Stemming	59
Gambar 4. 8 Hasil Penggabungan Judul Preprocessing	60
Gambar 4. 9 Grafik Pembobotan Pengujian 1	82
Gambar 4. 10 Grafik Evaluasi Pengujian 1	84
Gambar 4. 11 Grafik Evaluasi Pengujian 3	86
Gambar 4. 12 Hasil TF-IDF pada Judul S1	93
Gambar 4. 13 Hasil Kecocokan CBF untuk Judul S1	95
Gambar 4. 14 Hasil Probabilitas Kata untuk Judul S1	96
Gambar 4. 15 Hasil Probability Prior untuk Judul S1	97
Gambar 4. 16 Hasil Probability Posterior untuk Judul S1	98
Gambar 4. 17 Hasil Kecocokan NBC dengan Dosen untuk Judul S1	99
Gambar 4. 18 Hasil Pencocokan Akhir untuk Judul S1	100
Gambar 4. 19 Hasil TF-IDF untuk Judul S2	101
Gambar 4. 20 Hasil Rata-rata <i>Cosine Similarity</i> untuk Judul S1	102
Gambar 4. 21 Hasil Kecocokan CBF untuk Judul S2	102
Gambar 4. 22 Hasil Frekuensi Kata pada Judul S2	103
Gambar 4. 23 Hasil Probability Prior untuk Judul S2	104
Gambar 4. 24 Hasil Probability Posterior untuk Judul S2	105
Gambar 4. 25 Hasil Kecocokan NBC untuk Judul S2	106
Gambar 4. 26 Hasil Pencocokan Akhir untuk Judul S1	107

DAFTAR TABEL

Tabel 3. 1 Atribut Data Dosen	20
Tabel 3. 2 Jumlah Jurnal Setiap Dosen	21
Tabel 3. 3 Jumlah Jurnal Dosen Per-posisi	21
Tabel 3. 4 Data Skripsi Mahasiwa	22
Tabel 3. 5 Pemetaan Skripsi Mahasiswa.....	23
Tabel 3. 6 Contoh Hasil Translate.....	25
Tabel 3. 7 Contoh <i>Punctional Data</i>	26
Tabel 3. 8 Contoh <i>Case Folding</i>	26
Tabel 3. 9 Contoh <i>Tokenization</i>	27
Tabel 3. 10 Contoh <i>Stopword Removal</i>	28
Tabel 3. 11 Contoh <i>Stemming</i>	29
Tabel 3. 12 Contoh Frekuensi Kata dari Dokumen.....	31
Tabel 3. 13 Contoh Hasil DF dan IDF	32
Tabel 3. 14 Contoh Hasil TF-IDF	34
Tabel 3. 15 Contoh TF-IDF dari Q	35
Tabel 3. 16 Contoh Hasil <i>Cosine Similarity</i>	37
Tabel 3. 17 Contoh Data Jurnal Dosen	38
Tabel 3. 18 Contoh Jumlah Kata pada Dokumen Dosen	39
Tabel 3. 19 Contoh Hasil <i>Laplace Smoothing</i>	42
Tabel 3. 20 Skema Pembobotan Metode	43
Tabel 3. 21 Contoh Hasil dari Beberapa Skema	45
Tabel 3. 22 Table Pengujian.....	46
Tabel 4. 1 Tabel Skenario Pengujian	80
Tabel 4. 2 Tabel Data Pengujian 1	81
Tabel 4. 3 Data Pengujian 2	83
Tabel 4. 4 Data Pengujian 3	85
Tabel 4. 5 Data Uji Jurnal Dosen	88
Tabel 4. 6 Tabel Kesesuaian Jurnal J1 Dosen.....	89
Tabel 4. 7 Tabel Kesesuaian Jurnal J2 Dosen.....	89
Tabel 4. 8 Tabel Kesesuain Jurnal Dosen	90
Tabel 4. 9 Tabel Kesesuaian Jurnal J4 Dosen.....	91

Tabel 4. 10 Tabel Kesesuaian Jurnal J5 Dosen.....	91
Tabel 4. 11 Hasil Rata-rata Cosine Similarity pada Judul S1	94