



**PEMANFAATAN *TEXT MINING* DALAM PENENTUAN
REKOMENDASI DOSEN PEMBIMBING TUGAS AKHIR
PROGRAM STUDI SISTEM INFORMASI
(STUDI KASUS: UNIVERSITAS BINA DARMA)**

SKRIPSI

DEPTA RAHMA MULIA

181410288

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS TEKNOLOGI
UNIVERSITAS BINA DARMA
2023**



**PEMANFAATAN *TEXT MINING* DALAM PENENTUAN
REKOMENDASI DOSEN PEMBIMBING TUGAS AKHIR
PROGRAM STUDI SISTEM INFORMASI
(STUDI KASUS: UNIVERSITAS BINA DARMA)**

DEPTA RAHMA MULIA

181410288

**Skripsi ini diajukan sebagai syarat memperoleh gelar
Sarjana Komputer di Univeritas Bina Darma**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS TEKNOLOGI
UNIVERSITAS BINA DARMA
2023**

HALAMAN PENGESAHAN

**PEMANFAATAN *TEXT MINING* DALAM PENENTUAN
REKOMENDASI DOSEN PEMBIMBING TUGAS AKHIR PROGRAM
STUDI SISTEM INFORMASI
(STUDI KASUS: UNIVERSITAS BINA DARMA)**

DEPTA RAHMA MULIA

181410288

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi

Pembimbing

Palembang, 01 April 2023
Fakultas Sains Teknologi
Universitas Bina Darma
Dekan,



Universitas Bina Darma
Fakultas Sains Teknologi

Dr. Yesi Novaria Kunang, S.T., M.Kom Dr. Tata Sutabri, S.Kom., MMSI., MKM.

HALAMAN PERSETUJUAN

Skripsi berjudul "PEMANFAATAN *TEXT MINING* DALAM PENENTUAN REKOMENDASI DOSEN PEMBIMBING TUGAS AKHIR PROGRAM STUDI SISTEM INFORMASI (STUDI KASUS: UNIVERSITAS BINA DARMA)" oleh Depta Rahma Mulia telah dipertahankan didepan komisi penguji pada hari Sabtu tanggal 01 April 2023.

Komisi Penguji

1. Ketua : Dr. Yesi Novaria Kunang, S.T., M.Kom

(.....)

2. Anggota : Kurniawan, M.M., M.Kom

(.....)

3. Anggota : Suyanto, M.M., M.Kom

(.....)

Mengetahui,
Program Studi Sistem Informasi
Fakultas Sains Teknologi
Universitas Bina Darma
Ketua,



Nita Rosa Damayanti., M.Kom., Ph.D

SURAT PERANYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Depta Rahma Mulia

NIM : 181410288

Dengan ini menyatakan bahwa:

1. Karya tulis saya (Skripsi) adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya ;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing ;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan ;
4. Saya bersedia tugas skripsi, di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat diakses secara daring ;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang-undangan yang berlaku ;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 12 Mei 2023

Yang membuat pernyataan,



DEPTA RAHMA MULIA

NIM. 181410288

MOTTO DAN PERSEMBAHAN

Motto:

- ❖ **In the end, you only have God and Yourself to make everything you want come true.** (Pada akhirnya, kamu hanya memiliki Tuhan dan diri sendiri untuk mewujudkan segala sesuatu yang diinginkan.)
- ❖ Dibalik setiap kesulitan yang dialami akan ada indah diakhir cerita nanti.
- ❖ Sesungguhnya Allah tidak akan merubah suatu kaum sehingga mereka merubah keadaan yang ada pada diri mereka sendiri. (QS. Ar-Rad: 11)

Persembahan:

- ✚ Kedua orang tua yang aku cintai Bapak Rojali dan Ibu Supiyah yang selalu memberikan doa, nasehat, kasih sayang serta dukungan baik moral maupun material.
- ✚ Kakak Perempuanaku Siti Gia Hafiizha yang selalu memberikan dukungan yang aku minta dan selalu memberikan semangat.
- ✚ Dosen Pembimbingku Ibu Dr. Yesi Novaria Kunang, S.T., M.Kom, terima kasih atas bimbingan, nasehat, saran dan bantuan selama ini.
- ✚ Atas nama Fanatics (Angkic, Ayu, Dini, Ejak, Opik, dan Pika) terima kasih untuk dukungan yang telah kalian berikan.
- ✚ Teman seperjuangan disaat kuliah Social Squad dan Sebelah Besan Reborn terima kasih untuk 4 tahun di masa itu.
- ✚ Almamaterku

Abstract

The final assignment is the end of a student's journey in college where this final assignment is one of the determinants of whether or not students graduate. To be able to make the final assignment well, students get a supervisor who must be in accordance with the topic of the student's final assignment. Students studying at Bina Darma University will of course pass what is called a final project regardless of the major they take. In order to get supervisors who are in accordance with the topic of the student's final assignment, a basic system is created to recommend supervisors. This research was conducted using the BM25 method which is a method by means of ranking where the higher the BM25 score the lecturer's name will be at the top and it is most likely that the lecturer can be selected for students who input the topic of the final project. With raw data obtained from lecturer research data and student final assignment data several years ago, totaling 1473 data. From this data, it is then processed with text preprocessing and information retrieval. And then weighted with TF-IDF and finally, ranking is carried out by applying the program code from the BM25 score so that the results of the lecturer's recommendations are obtained.

Keywords : text mining, recommendation, final assignment, BM25

Abstrak

Tugas akhir adalah akhir dari perjalanan mahasiswa dalam masa kuliah yang mana tugas akhir ini menjadi salah satu penentu lulus tidaknya mahasiswa. Untuk dapat membuat tugas akhir dengan baik mahasiswa mendapat dosen pembimbing yang harus sesuai dengan topik tugas akhir dari mahasiswa tersebut. Mahasiswa yang berkuliah di Universitas Bina Darma tentunya akan melewati yang namanya tugas akhir dimanapun jurusan yang mereka ambil. Untuk mendapatkan dosen pembimbing yang memang sesuai dengan topik tugas akhir mahasiswa, dibuatlah suatu sistem dasar untuk merekomendasikan dosen pembimbing. Penelitian ini dilakukan dengan menggunakan metode BM25 yang merupakan sebuah metode dengan cara perankingan yang mana semakin tinggi skor BM25 tersebut nama dosen akan berada diatas dan kemungkinan besar dosen tersebut dapat terpilih bagi mahasiswa yang meng-*input* topik tugas akhir. Dengan data mentah yang didapat dari data penelitian dosen dan data tugas akhir mahasiswa beberapa tahun lalu yang berjumlah 1473 data. Dari data tersebut kemudian diproses dengan *text preprocessing* dan *information retrieval*. Setelah itu, dilakukan pembobotan dengan TF-IDF dan terakhir dilakukan perankingan dengan menerapkan kode program dari BM25 *score* sehingga terdapatlah hasil rekomendasi dosen.

Kata kunci: text mining, rekomendasi, tugas akhir, BM25

KATA PENGANTAR



Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-nyalah penulis dapat menyelesaikan skripsi yang berjudul “Pemanfaatan *Text Mining* dalam Penentuan Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi (Studi Kasus: Universitas Bina Darma)” dengan baik dan lancar.

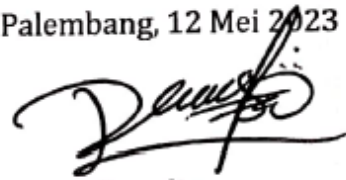
Tujuan dari penulisan skripsi ini adalah untuk syarat melanjutkan skripsi dan agar dapat menyelesaikan perkuliahan tepat pada waktunya. Pada kesempatan ini, penulis ingin menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan moril maupun materil sehingga skripsi ini dapat selesai. Ucapan terima kasih ini penulis tujukan kepada:

1. Allah SWT yang telah memberikan nikmat kesehatan & kesempatan sehingga penulis dapat menyelesaikan penelitian ini.
2. Kedua orang tua yang selalu memberikan dukungan kepada penulis.
3. Dr. Sunda Ariana, M.Pd., M.M selaku Rektor Universitas Bina Darma.
4. Dr. Tata Sutabri, S.Kom., MMSI, MKM selaku Dekan Fakultas Sains Teknologi.
5. Nita Rosa Damayanti, M.Kom., Ph.D. selaku Ketua Program Studi Sistem Informasi,
6. Dr. Yesi Novaria Kunang, ST., M.Kom selaku Dosen Pembimbing yang telah memberi bimbingan dalam penulisan penelitian ini.
7. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini, yang tidak bisa penulis sebutkan satu persatu.
8. Diri sendiri yang telah berjuang dalam penulisan skripsi ini baik dalam keadaan sehat maupun sakit.

Semoga Allah membalas semua kebaikan kalian dengan kebaikan yang lebih baik dan sebanyak-banyaknya, aamiin.

Akhir kata penulis berharap semoga penelitian ini dapat bermanfaat bagi kita semua.

Palembang, 12 Mei 2023



Penulis



DAFTAR ISI

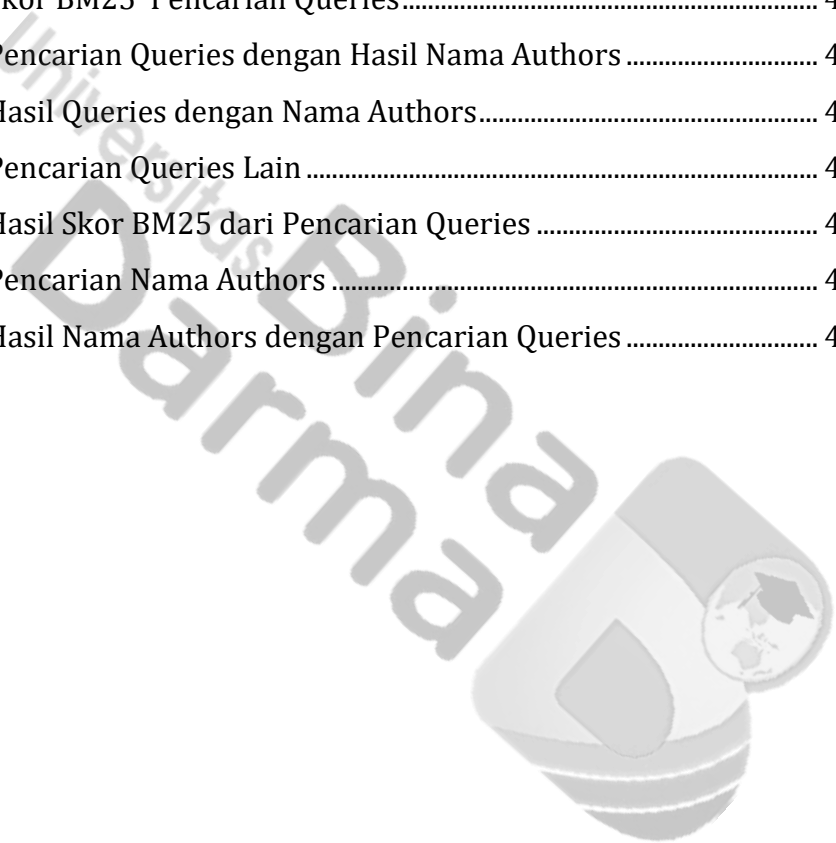
HALAMAN PENGESAHAN.....	ii
HALAMAN PERSETUJUAN	iii
SURAT PERANYATAAN	iv
MOTTO DAN PERSEMBAHAN.....	v
<i>Abstract</i>	vi
Abstrak	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	3
1.3 Rumusan Masalah	3
1.4 Batasan Masalah.....	3
1.5 Tujuan Penelitian	4
1.6 Manfaat Penelitian.....	4
1.7 Metodologi Penelitian.....	4
1.7.1 Waktu dan Tempat.....	4
1.7.2 Alat dan Bahan	4
1.7.3 Metode Penelitian	5
1.7.4 Metode Pengumpulan Data.....	5
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	7
2.1 TINJAUAN PUSTAKA.....	7
2.2 Dasar Teori	9
2.2.1 <i>Text Mining</i>	9
2.2.2 BM25	13
2.2.3 <i>Term Frequency – Inverse Document Frequency</i>	14

BAB III METODOLOGI PENELITIAN.....	16
3.1 Pengumpulan Data.....	17
3.1.1 <i>Crawling Data</i>	17
3.2 Preprocessing.....	19
3.2.1 <i>Case Folding</i>	19
3.2.2 <i>Remove Stopword</i>	20
3.2.3 <i>Tokenizing</i>	20
3.2.4 <i>Stemming</i>	20
3.3 Information Retrieval	20
3.4 Analisa.....	22
3.4.1 Analisa Data.....	22
3.4.2 Analisa Sistem.....	22
3.5 Transformation.....	22
3.5.1 Ekstraksi Fitur.....	22
3.5.2 Klasifikasi.....	23
3.5.3 Uji Model	23
3.5.4 Data Mining.....	24
3.5.5 <i>Evaluation</i> (Evaluasi)	24
BAB IV HASIL DAN PEMBAHASAN.....	25
4.1 Implementasi Proses.....	25
4.2 Proses Preprocessing.....	26
4.2.1 <i>Proses Case Folding</i>	26
4.2.2 <i>Proses Remove Stopword</i>	28
4.2.3 <i>Proses Tokenizing</i>	29
4.2.4 <i>Filtering dan Stemming</i>	29
4.3 Implementasi <i>Information Retrieval</i> dan Algoritma BM25	30
4.3.1 <i>Information Retrieval</i>	30
BAB V KESIMPULAN DAN SARAN.....	44
5.1 Kesimpulan.....	44
5.2 Saran	44
DAFTAR PUSTAKA.....	xv
LAMPIRAN.....	xviii

DAFTAR GAMBAR

Gambar 3. 1 Skema Penelitian.....	16
Gambar 3. 2 Halaman Awal Publish/Perish.....	17
Gambar 3. 3 Hasil Pencarian Penelitian Dosen dengan Publish	18
Gambar 3. 4 Data Mentah	18
Gambar 3. 5 Data Mentah Setelah Cleaning.....	19
Gambar 3. 6 Proses Information Retrieval.....	20
Gambar 4. 1 File Google Colaboratory.....	25
Gambar 4. 2 Import Packages.....	25
Gambar 4. 3 Kode Mengakses Data dari Google Drive.....	26
Gambar 4. 4 Kode Program dan Output dari Proses Pembacaan Data.....	26
Gambar 4. 5 Proses Feature Statistic	30
Gambar 4. 6 Hasil dari Feature Statistic	31
Gambar 4. 7 Proses dan Hasil dari Penghitungan Feature Statistic	31
Gambar 4. 8 Kode Program untuk menentukan Panjang Abstrak dan Title....	32
Gambar 4. 9 Grafik Panjang Abstrak dan Title	32
Gambar 4. 10 Kode Program Matplotlib untuk Visualisasi Grafik.....	32
Gambar 4. 11 Grafik Hasil.....	33
Gambar 4. 12 Proses Indexing Framework	33
Gambar 4. 13 Proses Pemvektoran TF-IDF.....	34
Gambar 4. 14 Proses Pembuatan Model TF-IDF	34
Gambar 4. 15 Hasil Model TF-IDF	34
Gambar 4. 16 Proses Pembuatan Fungsi Untuk BM25 Model	35
Gambar 4. 17 Pembobotan BM25 Default dan BM25 Specific.....	35
Gambar 4. 18 Hasil dari Pembobotan Model BM25 Default dan Specific	36
Gambar 4. 19 Query Formulation.....	36
Gambar 4. 20 Fungsi Retrieve_ranking.....	37
Gambar 4. 21 Fungsi Precision k.....	37
Gambar 4. 22 Fungsi F1-Score	37
Gambar 4. 23 Pembuatan Model TF-IDF, BM25 Default, dan BM25 Specific ..	38

Gambar 4. 24 Tabel Perbandingan Hasil Query TF-IDF, BM25 Default, dan BM25 Specific	38
Gambar 4. 25 Pembuatan Visualisasi Model TF-IDF, BM25 Default, dan BM25 Specific.....	38
Gambar 4. 26 Hasil Visualisasi	39
Gambar 4. 27 Peng-inputan Queries	39
Gambar 4. 28 Skor BM25 Pencarian Queries.....	40
Gambar 4. 29 Pencarian Queries dengan Hasil Nama Authors	40
Gambar 4. 30 Hasil Queries dengan Nama Authors.....	41
Gambar 4. 31 Pencarian Queries Lain	42
Gambar 4. 32 Hasil Skor BM25 dari Pencarian Queries	42
Gambar 4. 33 Pencarian Nama Authors	42
Gambar 4. 34 Hasil Nama Authors dengan Pencarian Queries	43



DAFTAR TABEL

Tabel 4. 1 Tabel Hasil Abstrak Sebelum dan Sesudah Cleaning.....	27
Tabel 4. 2 Tabel Hasil Abstrak Sebelum dan Sesudah Remove Stopwords ...	28
Tabel 4. 3 Tabel Abstrak Sebelum dan Sesudah Tokenizing.....	29
Tabel 4. 4 Tabel Abstrak Sebelum dan Sesudah Stemming.....	30

