



**ANALISIS FAKTOR PEMBOBOTAN ALGORITMA BM25 PADA PROSES
ANALISIS SENTIMEN PEMBELAJARAN TATAP MUKA**

ALISSA HAYATUNNUFUS

181410125

LAPORAN PENELITIAN

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS TEKNOLOGI

UNIVERSITAS BINA DARMA

2023



**ANALISIS FAKTOR PEMBOBOTAN ALGORITMA BM25 PADA PROSES
ANALISIS SENTIMEN PEMBELAJARAN TATAP MUKA**

ALISSA HAYATUNNUFUS

181410125

**Laporan Penelitian ini diajukan sebagai syarat memperoleh gelar
Sarjana Komputer**

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

HALAMAN PENGESAHAN

**Analisis faktor pembobotan algoritma BM25 pada
proses analisis sentimen pembelajaran tatap muka**

ALISSA HAYATUNNUFUS

181410125

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

Pembimbing

**Palembang, 11 September 2023
Fakultas Sains Teknologi
Universitas Bina Darma
Dekan,**



Dr. Yesi Novaria Kunang, S.T., M.Kom.




Universitas Bina Darma
Fakultas Sains Teknologi

Dr. Tata Sutabri, S.Kom., MMSI., MKM.

HALAMAN PERSETUJUAN

Skripsi Berjudul "Analisis faktor pembobotan algoritma BM25 pada proses analisis sentimen pembelajaran tatap muka" Oleh "Alissa Hayatunnufus", telah dipertahankan di depan komisi penguji pada hari Senin tanggal 11 September 2023.

Komisi Penguji

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

(.....)

2. Anggota : Heri Suroyo, M.Kom

(.....)

3. Anggota : Nyimas Sopiah, S.Kom., M.M., M.Kom.

(.....)

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

Universitas Bina
Darma
Fakultas Sains Teknologi

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

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Alissa Hayatunnufus

Nim : 181410125

Dengan ini menyatakan bahwa:

1. Karya tulis saya 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 karya tulis ini 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, September 2023

Yang membuat pernyataan



Alissa Hayatunnufus

NIM : 181410125

MOTTO DAN PERSEMBAHAN

MOTTO

- “Tidak ada kesuksesan tanpa kerja keras. Tidak ada keberhasilan tanpa kebersamaan. Tidak ada kemudahan tanpa doa.”

PERSEMBAHAN

Segala perjuangan saya hingga titik ini, saya persembahkan teruntuk orang-orang hebat yang selalu menjadi penyemangat, menjadi alasan saya kuat sehingga bisa menyelesaikan Karya Akhir ini.

1. Kedua orang tua tercinta saya ayah dan ibu yang sangat saya cintai dan sayangi, yang selalu memberikan kasih sayang, do'a, serta dukungan setiap langkah selama penulis menjalani masa perkuliahan, penulis berharap dapat menjadi anak yang dapat membanggakan keluarga. Beserta saudara perempuan dan laki-laki yang saya cintai dan sayangi.
2. Ibu Dr. Yesi Novaria Kunang, S.T., M.Kom selaku Dosen Pembimbing Laporan saya yang telah memberikan arahan dan bimbingan selama proses penelitian.
3. Ibu Nita Rosa Damayanti., M.Kom., Ph.D. selaku Ketua Program Studi Sistem Informasi Universitas Bina Darma.
4. Ibu Iin Seprina, M.Kom, selaku Dosen Pembimbing Akademik saya yang telah memberikan saran dan masukannya selama masa perkuliahan.
5. Kepada Anggun Pailina, Suci Rahma Puteri S.Kom dan Seseorang yang sudah selalu ada disaat penulis butuh hiburan dan kesulitan. Seluruh teman-teman dan pihak yang selalu *support* dan membantu saya dalam menyelesaikan tugas akhir ini.

ABSTRACT

The dissemination of information via social media triggers mixed reactions among users. The availability of information that is fast and easy to access gives rise to positive and negative opinions, as well as pros and cons among social media users. Twitter has an important role in conveying opinions about public figures, services and products. In this context, sentiment analysis is the key to understanding public opinion. This research focuses on weighting factors in sentiment analysis using the BM25 algorithm on opinions related to face-to-face learning. The research stages include data pre-processing, term weighting, and ranking using the BM25 algorithm. The BM25 algorithm has been proven effective in analyzing opinions by calculating the similarity of documents to queries. Other studies show that BM25 performs better than other weighting methods such as TF-IDF. The use of BM25 can help in information retrieval systems, such as in research related to Indonesian language sports news. This research aims to explore the weighting factors in sentiment analysis using the BM25 algorithm on opinions related to face-to-face learning. By understanding public sentiment, this research can provide deeper insight into the opinions that develop on social media platforms such as Twitter. The evaluation results of the K-Nearest Neighbor (KNN) model show an accuracy of 70.25%, with a high recall value for positive labels reaching 86%, while for negative labels it is 34%, and for neutral labels it is 46%. In terms of precision, this model reaches 76% for positive labels, 46% for negative labels, and 66% for neutral labels. These results were obtained through the use of 90% training data and 10% testing data, indicating the model's ability to classify data well in three different sentiment categories.

Keywords: Analysis, BM25, Opinion, Twitter

ABSTRAK

Penyebaran informasi melalui media sosial memicu reaksi beragam di kalangan pengguna. Ketersediaan informasi yang cepat dan mudah diakses memunculkan opini positif dan negatif, serta pro dan kontra di antara pengguna media sosial. Twitter memiliki peran penting dalam menyampaikan opini terhadap tokoh publik, layanan, dan produk. Dalam konteks ini, analisis sentimen menjadi kunci untuk memahami opini publik. Penelitian ini fokus pada faktor pembobotan dalam analisis sentimen menggunakan Algoritma BM25 pada opini terkait pembelajaran tatap muka. Tahapan penelitian meliputi pre-processing data, pembobotan term, dan pemeringkatan dengan algoritma BM25. Algoritma BM25 telah terbukti efektif dalam menganalisis opini dengan menghitung kemiripan dokumen terhadap query. Studi lain menunjukkan bahwa BM25 memiliki kinerja lebih baik daripada metode pembobotan lain seperti TF-IDF. Penggunaan BM25 dapat membantu dalam sistem temu kembali informasi, seperti pada penelitian terkait berita olahraga berbahasa Indonesia. Penelitian ini bertujuan untuk mendalami faktor pembobotan dalam analisis sentimen menggunakan Algoritma BM25 pada opini terkait pembelajaran tatap muka. Dengan memahami sentimen masyarakat, penelitian ini dapat memberikan wawasan lebih mendalam tentang opini yang berkembang di platform media sosial seperti Twitter. Hasil evaluasi model K-Nearest Neighbor (KNN) menunjukkan akurasi sebesar 70.25%, dengan nilai recall yang tinggi untuk label positif mencapai 86%, sementara untuk label negatif sebesar 34%, dan label netral sebesar 46%. Dalam hal precision, model ini mencapai 76% untuk label positif, 46% untuk label negatif, dan 66% untuk label netral. Hasil ini diperoleh melalui penggunaan data training sebesar 90% dan data testing sebesar 10%, mengindikasikan kemampuan model dalam mengklasifikasikan data dengan baik dalam tiga kategori sentimen yang berbeda.

Kata Kunci: Analisis, BM25, Opini, *Twitter*

KATA PENGANTAR

Puji dan syukur di panjatkan ke hadirat Allah swt. Yang Maha Kuasa, atas segala nikmat serta karuniaNya yang dilimpahkan kepada penulis, sehingga penulisan Skripsi ini dapat diselesaikan tepat waktu.

Skripsi ini berjudul “**Analisis Faktor Pembobotan Algoritma BM25 Pada Proses Analisis Sentimen Pembelajaran Tatap Muka**” Skripsi ini ditulis untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer (S.Kom.) di Program Studi Sistem Informasi Fakultas Sains Teknologi Universitas Bina Darma Palembang.

Penulisan Skripsi ini dapat diselesaikan dengan baik atas bantuan dari berbagai pihak. Dalam kesempatan ini penulis mengucapkan terima kasih yang sebesar besarnya kepada:

1. Dr. Sunda Ariana, M.Pd., M.M. Rektor Universitas Bina Darma Palembang;
2. Dr. Tata Sutabri, S.Kom., M.Si., MKM Dekan Fakultas Ilmu Komputer, Universitas Bina Darma Palembang;
3. Nita Rosa Damayanti., M.Kom., Ph.D. Ketua Program Studi Sistem Informasi Universitas Bina Darma Palembang;
4. Dr. Yesi Novaria Kunang, S.T., M.kom sebagai Pembimbing yang telah membimbing penulis dengan tulus dan sabar serta memberikan arahan sehingga penulisan Skripsi ini dapat diselesaikan;
5. Dosen Program Studi Sistem Informasi, Fakultas Sains Teknologi, Universitas Bina Darma yang telah mendidik serta memberikan ilmu pengetahuan;
6. Almamater Program Studi Sistem Informasi, Fakultas Sains Teknologi, Universitas Bina Darma Palembang, atas kesempatan yang diberikan untuk menimba ilmu di Lembaga ini.

Penulis berdoa kepada Allah swt semoga kebaikan yang mereka berikan kepada penulis mendapatkan balasan yang setimpal.

Semoga Skripsi ini dapat bermanfaat bagi pengembangan aplikasi dalam menunjang perkuliahan di Universitas.

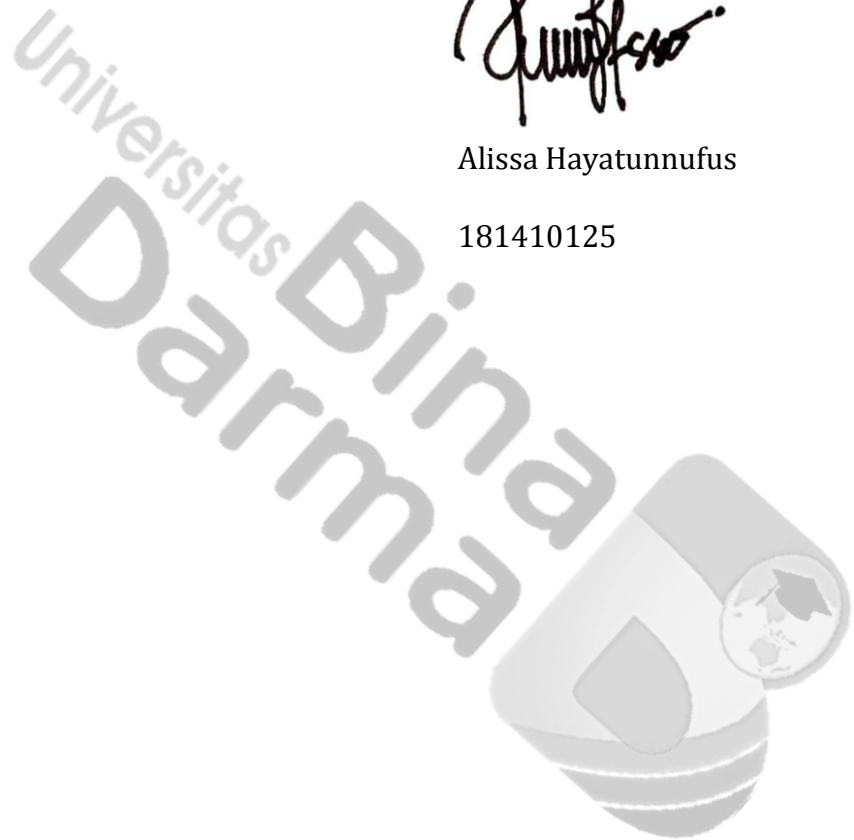
Palembang, Agustus 2023

Penulis,



Alissa Hayatunnufus

181410125



DAFTAR ISI

HALAMAN SAMPUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
SURAT PERNYATAAN	v
MOTTO DAN PERSEMBAHAN	vi
ABSTRACT.....	vii
ABSTRAK	viii
KATA PENGANTAR.....	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan dan Manfaat Penelitian.....	4
1.4.1 Tujuan Penelitian	4
1.4.2 Manfaat Penelitian.....	4
1.5 Sistematika penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Landasan Teori.....	7
2.1.1 Analisa	7
2.1.2 Analisa Sentimen.....	8
2.1.3 Klasifikasi.....	8
2.1.4 K-Nearest Neighbor	9
2.1.5 <i>Twitter</i>	9

2.1.6	<i>Preprocessing Text</i>	10
2.2	Algoritma BM25	11
2.3	Penelitian Sebelumnya	13
BAB III METODOLOGI PENELITIAN		15
3.1	Metode Pengumpulan Data	15
3.1.1	Studi Literatur	15
3.1.2	Observasi	15
3.2	Metode Penelitian	17
3.3	Tahapan Penelitian	18
3.4	Alat dan Bahan Penelitian	20
3.5	Proses Pengolahan Data	21
3.5.1	<i>Cleaning Data</i>	21
3.5.2	Preprocessing Data	22
3.5.3	<i>Tokenizing</i>	22
3.5.4	<i>Tranform Case</i>	23
3.5.5	<i>Stopword Removal</i>	24
3.5.6	Analisis Data	24
3.6	Hasil Analisis	25
BAB IV HASIL DAN PEMBAHASAN		26
4.1	Hasil Proses Pengumpulan Data	26
4.2	Hasil <i>Cleaning Data</i>	28
4.3	Hasil <i>Pre-Processing Data</i>	30
4.3.1	Fungsi Case Folding	31
4.3.2	Fungsi <i>Tokenize Data</i>	32
4.3.3	Fungsi <i>Stopword Removal</i>	34
4.3.3	Fungsi <i>Stemming</i>	37
4.4	Hasil Sentimen Analisis Data	38
4.4.1	Proses Pembobotan BM25	38
4.4.2	Hasil Pembagian Data	40
4.5	Pengujian Data Dengan <i>Python 3</i>	41

4.6 Hasil Pembahasan.....	45
BAB V PENUTUP	48
5.1 Kesimpulan.....	48
5.2 Saran.....	48
DAFTAR PUSTAKA.....	49



DAFTAR GAMBAR

Gambar 3. 1 Sampel Dataset.....	16
Gambar 3. 2 Tahapan Penelitian	19
Gambar 4. 1 Pemanggilan Fungsi <i>Read Pandas</i>	29
Gambar 4. 2 Hasil Fungsi <i>Read Excel</i>	30
Gambar 4. 3 Fungsi <i>Read Excel</i>	31
Gambar 4. 4 Fungsi <i>case folding</i>	31
Gambar 4. 5 Proses Fungsi <i>Tokenize</i>	32
Gambar 4. 6 Hasil Proses <i>Tokenize</i>	33
Gambar 4. 7 Fungsi <i>Stopword Removal</i>	35
Gambar 4. 8 Data Setelah Stopword Removal	36
Gambar 4. 9 Fungsi <i>Stemming</i>	37
Gambar 4. 10 Hasil Proses <i>Stemming</i>	38
Gambar 4. 11 Proses Pembobotan BM25	39
Gambar 4. 12 Hasil Nilai Bobot Query	39
Gambar 4. 13 Grafik Sentimen	41
Gambar 4. 14 Pemanggilan Fungsi Operator	42
Gambar 4. 15 Hasil Penarikan Dataset.....	42
Gambar 4. 16 Model <i>K-Nearest Neighbor (KNN)</i>	43
Gambar 4. 17 Fungsi Hasil Pengujian	43

DAFTAR TABEL

Tabel 2. 1 Penelitian Sebelumnya.....	13
Tabel 3. 1 Sampel Data Sentimen.....	16
Tabel 3. 2 Kebutuhan Perangkat Lunak.....	20
Tabel 3. 3 Kebutuhan Perangkat Keras.....	21
Tabel 3. 4 Proses <i>cleaning</i> data.....	22
Tabel 3. 5 Proses <i>Tokenizing</i>	23
Tabel 3. 6 Proses <i>tranform case</i>	24
Tabel 3. 7 Proses <i>Filtering</i>	24
Tabel 4. 1 Tabel Dataset Penelitian.....	26
Tabel 4. 2 Proses Perubahan Data Pada <i>Tokenize</i>	33
Tabel 4. 3 Proses Data <i>Stopword Removal</i>	36
Tabel 4. 4 Hasil Ranging Pembobotan BM25.....	40
Tabel 4. 5 Pembagian Data.....	40
Tabel 4. 6 Hasil Pengujian <i>Confussion Matrix</i>	44