

**ANALISIS SENTIMEN DAN TOPIK BERITA
TERKAIT PEMILU 2024 DI INDONESIA**

Tugas Akhir

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat
Sarjana Komputer**



Dibuat Oleh:

EMMANUEL MATHEW KRISNA RATA

200710530

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2024**

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

ANALISIS SENTIMEN DAN TOPIK BERITA TERKAIT PEMILU 2024 DI INDONESIA

yang disusun oleh

Emmanuel Mathew Krisna R

200710530

dinyatakan telah memenuhi syarat pada tanggal 24 Juli 2024

		Keterangan
Dosen Pembimbing 1	: Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D.	Telah Menyetujui
Dosen Pembimbing 2	: Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D.	Telah Menyetujui
Tim Penguji		
Penguji 1	: Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D.	Telah Menyetujui
Penguji 2	: Findra Kartika Sari Dewi, S.T., M.M., M.T.	Telah Menyetujui
Penguji 3	: Aloysius Gonzaga Pradnya Sidhawara, S.T., M.Eng.	Telah Menyetujui

Yogyakarta, 24 Juli 2024

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Dr. Ir. Parama Kartika Dewa SP., S.T., M.T.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Emmanuel Mathew Krisna Rata
NPM : 200701530
Program Studi : Informatika
Fakultas : Teknologi Industri
Judul Penelitian : ANALISIS SENTIMEN DAN TOPIK BERITA
TERKAIT PEMILU 2024 DI INDONESIA

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 24 Juli 2024

Yang menyatakan,

Emmanuel Mathew Krisna Rata

200710530

HALAMAN PERSEMBAHAN

“Selalu tekun ,sabar dan usaha di setiap ujian yang di berikan pasti menghasilkan buah yang baik”

KATA PENGANTAR

Puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan pembuatan tugas akhir “Analisis Sentimen dan Topik Berita Terkait Pemilu 2024 di Indonesia” ini dengan baik.

Penulisan tugas akhir ini bertujuan untuk memenuhi salah satu syarat untuk mencapai derajat sarjana komputer dari Program Studi Informatika, Fakultas Teknologi Industri di Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan tugas akhir ini penulis telah mendapatkan bantuan, bimbingan, dan dorongan dari banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Yesus Kristus yang selalu membimbing dalam iman-Nya, memberikan berkat-Nya, dan menyertai penulis selalu.
2. Bapak Dr. Ir. Parama Kartika Dewa SP., S.T., M.T., IPU. , selaku Dekan Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
3. Bapak Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D., selaku dosen pembimbing I yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
4. Bapak Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D. selaku dosen pembimbing II yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
5. Kepada Kedua orang tua yang selalu mendoakan dan memberi dukungan baik moral dan material kepada penulis.
6. Kepada Para sahabat dan teman-teman yang selalu mendoakan dan menyemangati penulis.
7. Kepada Semua pihak yang tidak dapat disebutkan satu per satu yang selalu

memberikan doa, dukungan, dan semangat kepada penulis.

Demikian laporan tugas akhir ini dibuat, dan penulis mengucapkan terima kasih kepada semua pihak. Semoga laporan ini dapat bermanfaat bagi pembaca.

Yogyakarta, 24 Juli 2024

Emmanuel Mathew Krisna Rata

200701530

DAFTAR ISI

ANALISIS SENTIMEN DAN TOPIK BERITA TERKAIT PEMILU 2024 DI INDONESIA	i
LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
INTISARI	xii
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan Masalah	2
C. Batasan Masalah	2
D. Tujuan Penelitian	3
E. Metode Penelitian	3
F. Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI	13
1. Web Crawling	13
2. <i>Preprocessing</i>	13
3. Analisis Sentimen	14
4. Klasifikasi	15
5. Pemodelan Topik	17
6. Grid Search	18
7. Metrik Evaluasi	19
BAB IV METODOLOGI PENELITIAN	20
A. Proses Pengumpulan Data	21

B. <i>Pre-processing</i> Data	21
C. Pemrosesan Model	22
D. Implementasi Model	25
BAB V HASIL DAN PEMBAHASAN	27
A. Proses Pengumpulan Data	27
B. <i>Pre-processing</i> Data	32
C. Pemrosesan Model	35
D. Implementasi Model	67
BAB VI PENUTUP	71
A. Kesimpulan	71
B. Saran	72
DAFTAR PUSTAKA	73

DAFTAR GAMBAR

Gambar 4. 1 Gambar Skema Metodologi Penelitian	20
Gambar 4. 2 Gambar Skema Crawling Berita	21
Gambar 5. 1 Gambar Hasil Crawling Berita di Database	27
Gambar 5. 2 Gambar Fungsi Pengambilan Link Pada Detik	28
Gambar 5. 3 Gambar Kode Crawling Data pada Detik	29
Gambar 5. 4 Gambar Kode Proses Memasukan kedalam Database	30
Gambar 5. 5 Gambar Terminal Proses Pengambilan Link Berita	31
Gambar 5. 6 Gambar Terminal Proses Melakukan Pengambilan Data	31
Gambar 5. 7 Gambar Kode Filter Data	31
Gambar 5. 8 Gambar Hasil Filter Data	32
Gambar 5. 9 Gambar Kode Pembersihan Data	32
Gambar 5. 10 Gambar Kode Case Folding	33
Gambar 5. 11 Gambar Kode Stopword Removal	34
Gambar 5. 12 Gambar Kode Tokenisasi	35
Gambar 5. 13 Gambar Dataset Pelabelan 3.000 Data	36
Gambar 5. 14 Gambar Kode Filter Ukuran Content	37
Gambar 5. 15 Gambar Kode Otomatis Pelabelan	37
Gambar 5. 16 Gambar Proses Pelabelan Otomatis	38
Gambar 5. 17 Gambar Dataset Pelabelan 12.282 Data	38
Gambar 5. 18 Gambar Kode Metode TF-IDF, Bag Of Word (Bow)	39
Gambar 5. 19 Gambar Diagram Batang 1500 Dataset Manual	40
Gambar 5. 20 Gambar Word Cloud Positif Dataset 1500 Manual	41
Gambar 5. 21 Gambar Word Cloud Netral Dataset 1500 Manual	41
Gambar 5. 22 Gambar Word Cloud Negatif Dataset 1500 Manual	41
Gambar 5. 23 Gambar Diagram Batang 1500 Dataset Otomatis	42
Gambar 5. 24 Gambar Word Cloud Positif Dataset 1500 Otomatis	43
Gambar 5. 25 Gambar Word Cloud Netral Dataset 1500 Otomatis	43
Gambar 5. 26 Gambar Word Cloud Negatif Dataset 1500 Otomatis	43
Gambar 5. 27 Gambar Diagram Batang 2400 Dataset Manual	45
Gambar 5. 28 Gambar Word Cloud Positif Dataset 2400 Manual	46

Gambar 5. 29 Gambar Word Cloud Netral Dataset 2400 Manual	46
Gambar 5. 30 Gambar Wordl Cloud Netral Dataset 2400 Manual	46
Gambar 5. 31 Gambar Diagram Batang 2400 Dataset Otomatis	47
Gambar 5. 32 Gambar Word Cloud Positif Dataset 2400 Otomatis	48
Gambar 5. 33 Gambar Word Cloud Netral Dataset 2400 Otomatis	48
Gambar 5. 34 Gambar Word Cloud Negatif Dataset 2400 Otomatis	48
Gambar 5. 35 Gambar Diagram Batang 3000 Dataset Manual	50
Gambar 5. 36 Gambar Word cloud Positif Dataset 3000 Manual	51
Gambar 5. 37 Gambar Word Cloud Netral Dataset 3000 Manual	51
Gambar 5. 38 Gambar Word cloud Negatif Dataset 3000 Manual	52
Gambar 5. 39 Gambar Diagram Batang 3000 Dataset Otomatis	53
Gambar 5. 40 Gambar Word Cloud Positif Dataset 3000 Otomatis	53
Gambar 5. 41 Gambar Word cloud Netral Dataset 3000 Otomatis	54
Gambar 5. 42 Gambar Word Cloud Netral Dataset 3000 Otomatis	54
Gambar 5. 43 Gambar Diagram Batang Perbandingan Dataset	56
Gambar 5. 44 Gambar Kode Probilatas Sentimen	58
Gambar 5. 45 Gambar Kode Pengecek Nilai Coherence Score	60
Gambar 5. 46 Gambar Diagram Garis Coherence Score LDA	61
Gambar 5. 47 Gambar Kode LDA	62
Gambar 5. 48 Gambar Diagram Garis Coherence Score LSA	63
Gambar 5. 49 Gambar Kode LSA	64
Gambar 5. 50 Gambar World Cloud LDA	65
Gambar 5. 51 Gambar Bar Chart LDA	65
Gambar 5. 52 Gambar Worldcloud LSA	66
Gambar 5. 53 Gambar Bar Chart LSA	66
Gambar 5. 54 Gambar Cuplikan Baranda Statistik Dataset	67
Gambar 5. 55 Gambar Cuplikan Baranda Statistik Senitmen	68
Gambar 5. 56 Cuplikan gambar Prediksi Senitmen	69
Gambar 5. 57 Gambar Cuplikan LDA	70
Gambar 5. 58 Gambar Cuplikan LSA	70

DAFTAR TABEL

Tabel 2. 1 Tabel Perbandingan Tinjauan Pustaka	12
Tabel 4. 1 Tabel Penilaian Sentimen	22
Tabel 4. 2 Tabel Contoh Labeling Otomatis	23
Tabel 5. 1 Tabel Pembersihan Data	33
Tabel 5. 2 Tabel Case Folding	34
Tabel 5. 3 Tabel Stopword Removal	34
Tabel 5. 4 Tabel Tokenisasi	35
Tabel 5. 5 Tabel 1500 Dataset Manual	40
Tabel 5. 6 Tabel 1500 Dataset Otomatis	42
Tabel 5. 7 Tabel 2400 Dataset Manual	45
Tabel 5. 8 Tabel 2400 Dataset Otomatis	47
Tabel 5. 9 Tabel 3000 Dataset Manual	50
Tabel 5. 10 Tabel 3000 Dataset Otomatis	52
Tabel 5. 11 Tabel Perbandingan Dataset	55
Tabel 5. 12 Tabel Coherence Score Topik LDA	61
Tabel 5. 13 Tabel Topik LDA	62
Tabel 5. 14 Tabel Coherence Score Top Topik LSA	63
Tabel 5. 15 Tabel Topik LSA	64

INTISARI

ANALISIS SENTIMEN DAN TOPIK BERITA TERKAIT PEMILU 2024 DI INDONESIA

Intisari

Emmanuel Mathew Krisna Rata

200710530

Pemilihan umum (Pemilu) di Indonesia merupakan bentuk penerapan demokrasi yang melibatkan partisipasi aktif seluruh warga negara dalam proses pemilihan yang adil dan bebas serta pemberdayaan rakyat dalam pengambilan keputusan politik dan sosial. Demokrasi menjamin bahwa setiap suara didengar melalui dialog terbuka dan kebebasan berekspresi. Menjelang Pemilu 2024, peran media, baik cetak maupun elektronik, sangat penting dalam memberikan informasi dan opini kepada masyarakat yang berpotensi mempengaruhi pandangan terhadap partai politik dan calon pemimpin. Analisis sentimen menjadi metode untuk mengevaluasi pendapat publik tentang suatu topik, baik secara kuantitatif menggunakan statistik maupun kualitatif dengan teknik interpretasi. Dalam penelitian ini, penggunaan algoritme yang tepat untuk analisis sentimen dan topik sangat penting untuk memperoleh hasil dengan akurasi tinggi, dengan mempertimbangkan efisiensi model dari penelitian sebelumnya. Adapun masalah yang dibahas meliputi metode crawling portal berita online, model klasifikasi untuk sentimen isi berita dengan membandingkan beberapa algoritme klasifikasi, hasil topik dominan dari isi berita menggunakan model topik *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA), serta implementasi hasil model sentimen dan pemodelan topik.

Masalah-masalah yang diangkat, memiliki metode untuk menyelesaikan masalah tersebut. Metodologi penelitian yang dimulai dengan pengumpulan data melalui metode crawling, kemudian data di-filter untuk kata-kata terkait pemilu dan diproses melalui tahap *pre-processing*. Selanjutnya, data diproses dalam dua bagian: pertama, proses klasifikasi yang meliputi pelabelan data secara manual dan otomatis pada isi berita, diikuti dengan proses klasifikasi dan perbandingan algoritme; kedua, pemodelan topik di mana isi berita digunakan untuk membuat topik LDA dan LSA serta divisualisasikan. Terakhir, implementasi dari hasil klasifikasi dan pemodelan topik dilakukan menggunakan Streamlit.

Proses crawling berita Indonesia dilakukan dengan mengambil berita harian berdasarkan tanggal tertentu, kemudian dilanjutkan dengan pengambilan data URL yang disimpan ke dalam database untuk mendapatkan berita nasional setiap harinya. Model sentimen yang digunakan untuk klasifikasi menunjukkan

bahwa kombinasi TF-IDF dan Bag of Words (BoW) pada dengan algoritme terbaik jatuh di Support Vector Machine (SVM) menghasilkan akurasi tertinggi sebesar 90.89% dan F1 skor 90.86%. Akurasi ini dipengaruhi oleh kualitas dataset yang digunakan. Pada pemodelan topik, LDA memiliki coherence score 0.6089501917687159 dengan topik utama meliputi kata gibran, prabowo, ganjar, debat, dan anies, sedangkan LSA memiliki coherence score 0.5654880904606634 dengan satu topik utama yang mencakup kata gibran, prabowo, ganjar, cawapres, dan presiden. Hasil implementasi model sentimen dan pemodelan topik tersebut di deploy ke dalam Streamlit dengan tiga menu: beranda yang menjelaskan tentang dataset yang digunakan, model klasifikasi yang menerapkan model sentimen, dan model pemodelan topik yang menampilkan hasil dari LDA dan LSA, yang semuanya dapat diakses melalui URL ta10530.streamlit.app.

Kata Kunci: crawling, pemodelan topik, sentimen, algoritme klasifikasi

Dosen Pembimbing I : Y. Sigit Purnomo WP., S.T., M.Kom., Ph.D.

Dosen Pembimbing II : Prof. Ir. A. Djoko Budiyanto, M.Eng., Ph.D.

Jadwal Sidang Tugas Akhir : 17 juli 2024

BAB I

PENDAHULUAN

A. Latar Belakang

Pemilihan umum (Pemilu) tiada lain merupakan penerapan demokrasi di Indonesia. berbagai langkah dan upaya yang melibatkan partisipasi aktif dari seluruh warga negara. Proses ini melibatkan pemilihan umum yang adil dan bebas, serta pemberdayaan rakyat dalam mengambil keputusan politik dan sosial yang berpengaruh terhadap kehidupan mereka. Melalui dialog yang terbuka dan kebebasan berekspresi, demokrasi memastikan bahwa setiap suara didengar dan dihargai [1].

Menjelang Pemilu 2024, peran media baik cetak maupun elektronik sangat penting dalam memberikan informasi dan opini kepada masyarakat. Informasi yang dimuat di media massa berpotensi mempengaruhi pandangan masyarakat terhadap partai politik dan calon pemimpinnya [6].

Analisis sentimen adalah metode untuk mengevaluasi pendapat publik dan perspektif masyarakat tentang suatu topik. Dua metode berbeda dapat digunakan untuk melakukan analisis sentimen: metode kuantitatif menggunakan statistik untuk mengukur jumlah emosi positif, negatif, atau netral yang ada dalam teks; metode kualitatif menggunakan teknik interpretasi untuk memahami makna teks dan menafsirkannya.

Dalam menganalisis sentimen dan topik menggunakan algoritme. algoritme yang tepat dapat sentimen dan topik dengan hasil yang akurasi tinggi, maka dari algoritme yang di pakai dalam penelitian ini adalah algoritme yang sudah memiliki model yang sudah jauh efisien dari penelitian sebelumnya yang menjadi bahan pertimbangan dan lebih membahas lain untuk mengoptimalkan model dari algoritma sebelumnya.