

**ANALISIS SENTIMEN MULTIBAHASA DENGAN PELABELAN ZERO
SHOT PADA BERITA PEMILU 2024**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

JODI ABDILLAH

19.11.2961

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

**ANALISIS SENTIMEN MULTIBAHASA DENGAN PELABELAN ZERO
SHOT PADA BERITA PEMILU 2024**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

JODI ABDILLAH

19.11.2961

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN MULTIBAHASA DENGAN PELABELAN ZERO
SHOT PADA BERITA PEMILU 2024**

yang disusun dan diajukan oleh

Jodi Abdillah

19.11.2961

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Agustus 2024

Dosen Pembimbing,



Hastari Utama, M.Cs

NIK. 190302230

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN MULTIBAHASA DENGAN PELABELAN ZERO
SHOT PADA BERITA PEMILU 2024

yang disusun dan diajukan oleh

Jodi Abdillah

19.11.2961

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Agustus 2024

Susunan Dewan Penguji

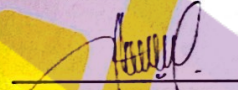
Nama Penguji

Tanda Tangan

Lukman, S.Kom., M.Kom
NIK. 190302151



Mulia Sulistiyono, M.Kom
NIK. 190302248



Hastari Utama, M.Cs
NIK. 190302230



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Agustus 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Jodi Abdillah
NIM : 19.11.2961

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS SENTIMEN MULTIBAHASA DENGAN PELABELAN ZERO SHOT PADA BERITA PEMILU 2024

Dosen Pembimbing : Hastari Utama, M.Cs

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 12 Agustus 2024

Yang Menyatakan,



Jodi Abdillah

HALAMAN PERSEMBAHAN

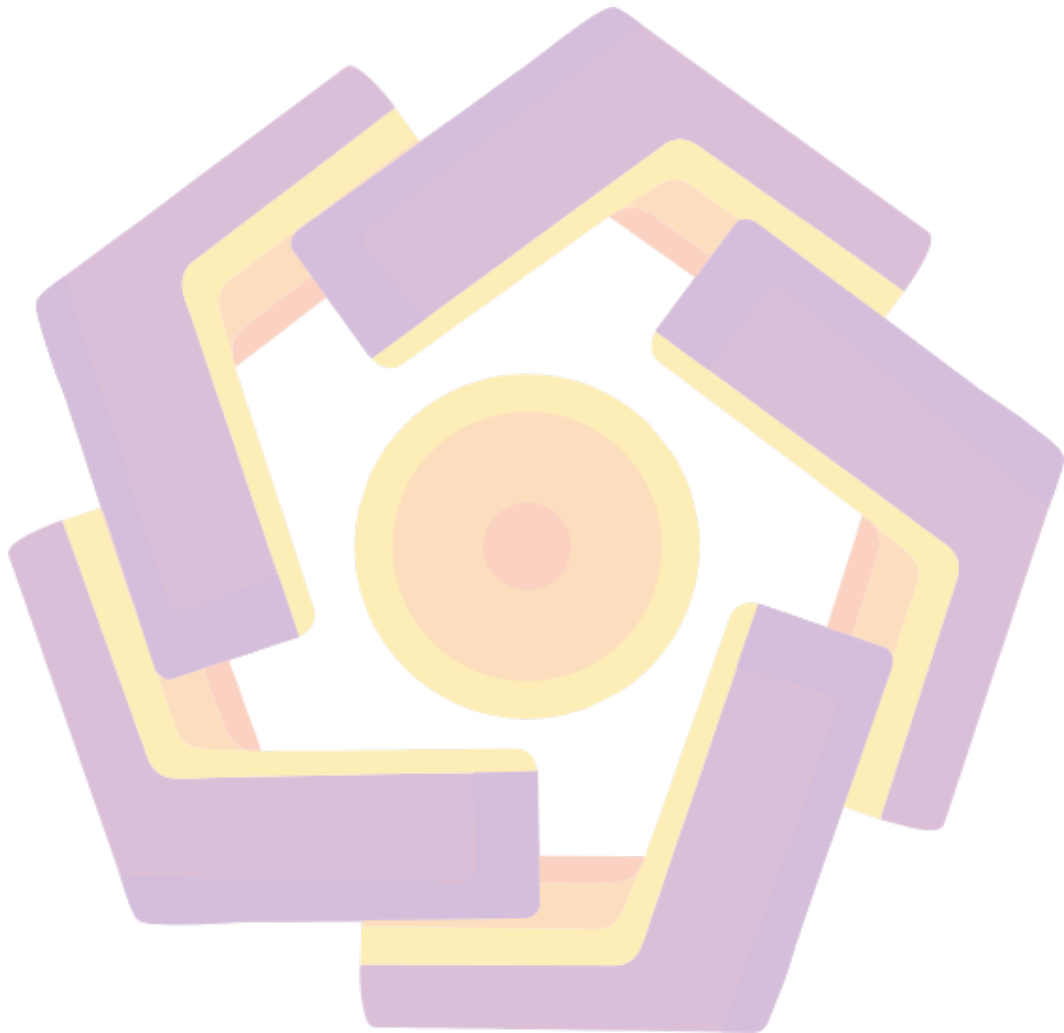
Puji dan syukur kehadirat Allah SWT atas limpahan berkat dan rahmat-Nya sehingga Skripsi ini dapat terselesaikan dengan baik. Terimakasih juga saya haturkan kepada seluruh pihak yang telah membantu dan memberikan dukungan dalam penyusunan Skripsi ini.

Pada Skripsi ini dimungkinkan masih banyak kekurangan yang harus diperbaiki. Selain itu, juga terdapat hasil yang kurang maksimal sehingga segala bentuk kritik dan saran akan diterima dengan senang hati dan diharapkan mampu membantu dalam penulisan laporan selanjutnya agar lebih baik lagi. Semoga Skripsi ini dapat menambah wawasan dan pengetahuan bagi pembaca.

Adapun penyusunan Skripsi digunakan sebagai bukti bahwa penyusun telah melakukan dan menyelesaikan penelitian Skripsi tugas akhir. Dalam proses penyusunan laporan ini penulis mendapatkan banyak bantuan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D. selaku dekan fakultas ilmu komputer.
3. Ibu Windha Mega Pradnya Dhuhita, M.Kom. selaku ketua program studi.
4. Bapak Hastari Utama, M.Cs. selaku dosen pembimbing
5. Bapak Arifiyanto Hadinegoro, S.T., M.Kom., Bapak Subektiningsih, M.Kom, Bapak M. Tofa Nurholis, M.Kom., Ibu Anna Baita M.Kom, Bapak M. Anul Fikri, S.Kom M.Eng, Bapak Bayu Setiaji, M. Kom, Theopilus Bayu Sasongko, M.Eng., Ibu Nuri Cahyono, M.Kom, dan Bapak Uyock Anggoro Saputro, M.Kom. selaku dosen – dosen program studi informatika.
6. Bapak Joko Ahmad Siswanto selaku orang tua penulis.

7. Sonya Qonita selaku pasangan penulis.
8. Seluruh kerabat, saudara dan rekan – rekan yang mendoakan serta mendukung terselesaikannya Skripsi ini.



KATA PENGANTAR

Puji Syukur bagi Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul “ANALISIS SENTIMEN MULTIBAHASA DENGAN PELABELAN ZERO SHOT PADA BERITA PEMILU 2024” ini dengan baik. Selama proses penulisan skripsi ini, penulis banyak mendapatkan bantuan, pengarahan, dan bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini sebagai wujud rasa hormat, penulis mengucapkan terima kasih yang sebesar besarnya :

- 1 Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
- 2 Ibu Windha Mega Pradnya D, M.Kom selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
- 3 Bapak Hastari Utama, M.Cs selaku dosen pembimbing yang selalu memberi saran dan masukan dalam penulisan skripsi ini.
- 4 Bapak Lukman, S.Kom. dan Bapak Mulia Sulistiyono, M.Kom. selaku dosen penguji yang telah memberi saran dan masukan dalam penulisan skripsi ini.
- 5 Bapak Joko Ahmad Siswanto ayah dari keluarga penulis yang senantiasa mendoakan dan memberikan dukungan kepada penulis.
- 6 Semua pihak termasuk pasangan, teman dan saudara yang senantiasa menjadi penyemangat dan menghibur penulis sehingga skripsi ini dapat selesai dengan baik.

Penulis menyadari bahwa skripsi ini masih terdapat banyak kekurangan baik dalam isi maupun cara penyajian. Oleh karena itu, penulis mengharapkan kritik dan saran guna perbaikan di masa mendatang.

Yogyakarta, 12 Agustus 2024



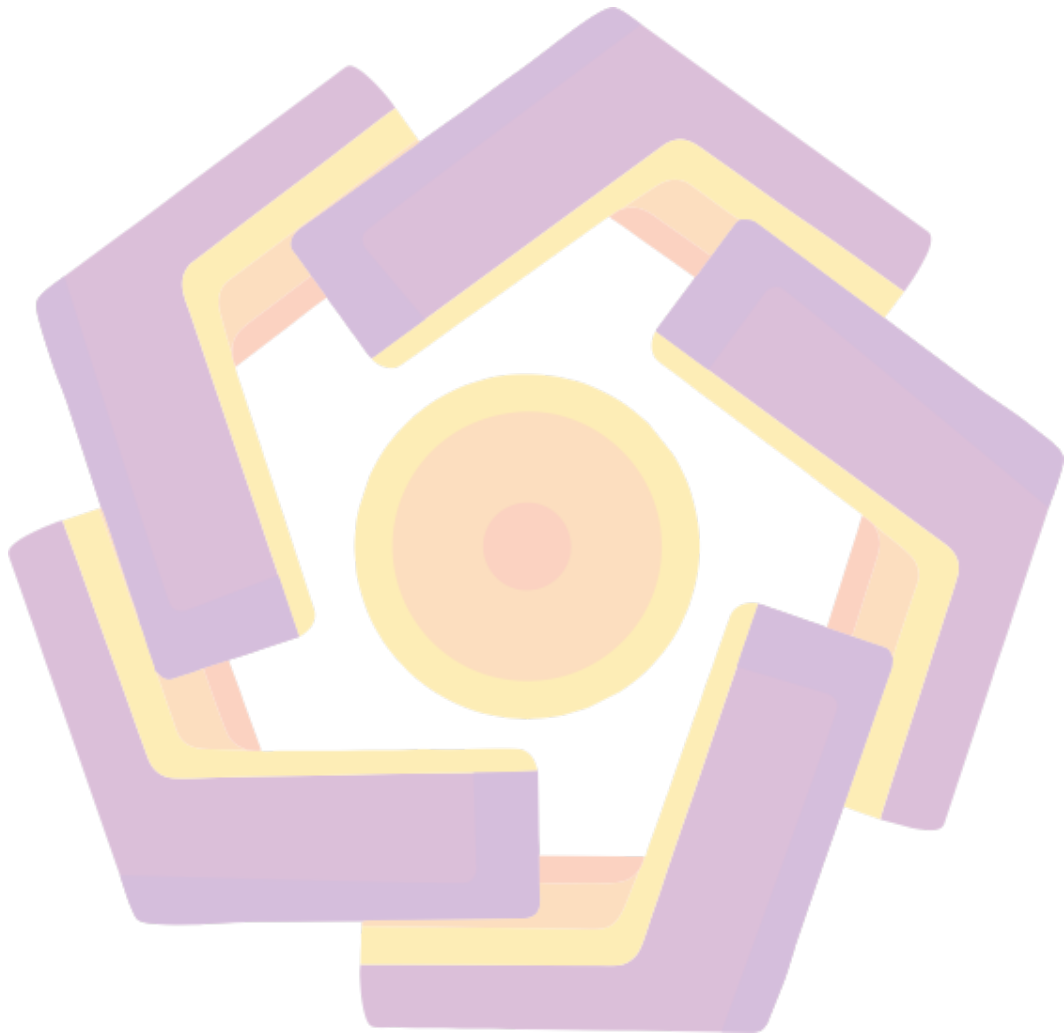
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR	vii
DAFTAR ISI.....	1
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi <i>Literatur</i>	6
2.2 Berita	15
2.3 Pemilu.....	15
2.4 <i>Data Mining</i>	16
2.4.1 Pengertian <i>Data Mining</i>	16
2.4.2 Kegunaan <i>Data Mining</i>	16
2.5 Analisis Sentimen	17
2.6 <i>Scraping Website</i>	18
2.7 <i>Beautiful Soup</i>	18

2.8	<i>Text Mining</i>	18
2.8.1	Akuisisi Data.....	19
2.8.2	<i>Text Pre-processing</i>	19
2.9	Klasifikasi.....	20
2.10	Algoritma <i>Support Vector Machine (SVM)</i>	20
2.11	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	23
2.12	TF-IDF.....	23
2.13	Google Trans.....	24
2.14	<i>Zero-shot Learning</i>	25
2.15	Evaluasi.....	26
BAB III METODE PENELITIAN		30
3.1	Objek Penelitian.....	30
3.2	Metode Pengumpulan Data.....	30
3.3	Alur Penelitian.....	30
3.3.1	Analisis Sistem.....	30
3.3.2	Perancangan Sistem.....	33
3.4	Alat dan Bahan Penelitian.....	47
3.4.1	Perangkat Keras.....	47
3.4.2	Perangkat Keras.....	47
BAB IV HASIL DAN PEMBAHASAN		48
4.1	Implementasi Pengambilan <i>Dataset</i>	48
4.2	Implementasi <i>Translate Data</i>	48
4.3	Implementasi <i>Labelling Data</i>	53
4.4	Implementasi <i>Pre-processing Data</i>	54
4.5	Implementasi <i>TF-IDF Vectorizer</i>	62
4.6	Implementasi <i>Split data dan Oversampling Dataset</i>	67
4.7	Implementasi Algoritma <i>Support Vector Machine (SVM)</i>	69
4.8	Hasil dan Evaluasi.....	70
4.8.1	Hasil.....	70
4.8.2	Evaluasi.....	72
BAB V PENUTUP		84

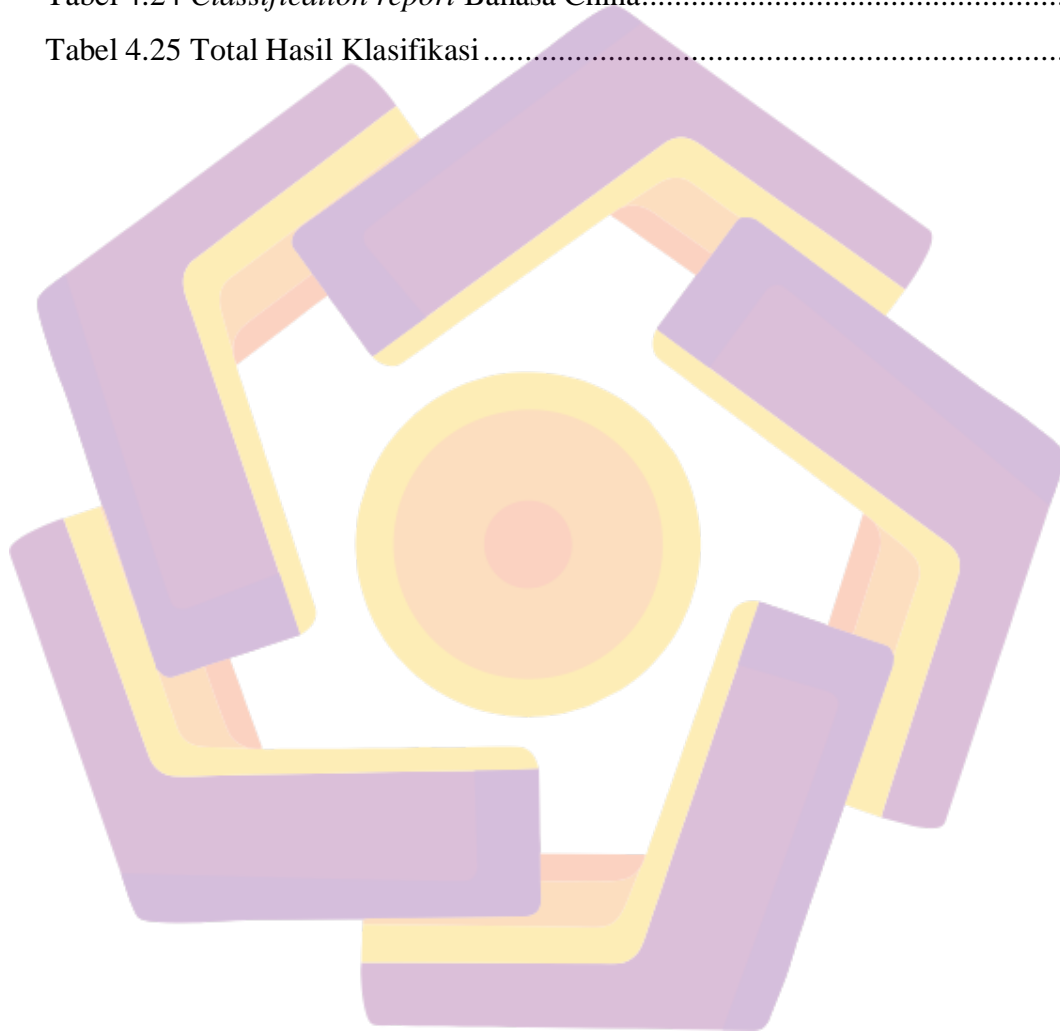
5.1 Kesimpulan.....	84
5.2 Saran.....	85
REFERENSI.....	86



DAFTAR TABEL

Tabel 2.1 <i>Studi Literatur</i> Penelitian Terdahulu.....	10
Tabel 3.1 Contoh <i>Split Text</i>	38
Tabel 3.2 Contoh <i>Remove Tag</i>	38
Tabel 3.3 Contoh <i>Remove Tag To Sentence</i>	39
Tabel 3.4 Contoh <i>Stemming Text To Sentence</i>	40
Tabel 3.5 Contoh <i>Case Folding</i>	40
Tabel 3.6 Contoh <i>CLearning Text</i>	41
Tabel 3.7 Contoh <i>Tokenization</i>	41
Tabel 3.8 Contoh <i>Stopword Removal</i>	42
Tabel 3.9 Contoh <i>Stemming Text</i>	43
Tabel 4.1 Hasil <i>Translate Data</i>	49
Tabel 4.2 Hasil <i>Labelling Data</i>	53
Tabel 4.3 Contoh Hasil <i>Split data</i>	54
Tabel 4.4 Contoh Hasil <i>Remove Tag</i>	55
Tabel 4.5 Contoh Hasil <i>To Sentence</i>	56
Tabel 4.6 Contoh Hasil <i>Case Folding</i>	57
Tabel 4.7 Contoh Hasil <i>Cleaning Text</i>	58
Tabel 4.8 Contoh Hasil <i>Tokenization</i>	59
Tabel 4.9 Contoh Hasil <i>Stopword Removal</i>	60
Tabel 4.10 Contoh Hasil <i>Stemming</i>	61
Tabel 4.11 Contoh Hasil <i>To Sentence</i>	62
Tabel 4.12 Contoh Hasil Perhitungan BOW.....	63
Tabel 4.13 Rincian Hasil Perhitungan TF-IDF.....	66
Tabel 4.14 Hasil Akurasi Setiap Bahasa.....	70
Tabel 4.15 <i>Classification Report</i> Bahasa Arab.....	72
Tabel 4.16 <i>Classification Report</i> Bahasa Bulgaria.....	72
Tabel 4.17 <i>Classification Report</i> Bahasa Jerman.....	73
Tabel 4.18 <i>Classification Report</i> Bahasa Yunani.....	73

Tabel 4.19 <i>Classification Report</i> Bahasa Inggris	73
Tabel 4.20 <i>Classification Report</i> Bahasa Spanyol	74
Tabel 4.21 <i>Classification Report</i> Bahasa Prancis.....	75
Tabel 4.22 <i>Classification Report</i> Bahasa Rusia	75
Tabel 4.23 <i>Classification Report</i> Bahasa Turki	75
Tabel 4.24 <i>Classification report</i> Bahasa China.....	76
Tabel 4.25 Total Hasil Klasifikasi	76



DAFTAR GAMBAR

Gambar 2.1 Contoh <i>Confusion Matrix</i>	27
Gambar 3.1 Alur Sistem Analisis Sentimen	31
Gambar 3.2 Alur Proses <i>Labelling</i> Dengan <i>Zero shot</i>	35
Gambar 3.3 Alur <i>Pre-processing</i>	37
Gambar 3.4 Alur Evaluasi.....	46
Gambar 4.1 Hasil <i>Vocabulary</i> BOW	63
Gambar 4.2 Hasil Representasi BOW	63
Gambar 4.3 Contoh Hasil Perhitungan TF.....	65
Gambar 4.4 Contoh Hasil Perhitungan IDF.....	65
Gambar 4.5 Contoh Hasil Perhitungan IDF.....	66
Gambar 4.6 Baris Kode <i>Split data</i>	67
Gambar 4.7 Diagram Perbandingan Data SMOTE	68
Gambar 4.8 Baris Kode <i>Modelling</i> SVM	69
Gambar 4.9 Grafik Uji Analisis Sentimen	71
Gambar 4.10 <i>Confusion Matrix</i> Bahasa Arab	77
Gambar 4.11 <i>Confusion Matrix</i> Bahasa Bulgaria.....	78
Gambar 4.12 <i>Confusion Matrix</i> Bahasa Jerman.....	78
Gambar 4.13 <i>Confusion Matrix</i> Bahasa Yunani	79
Gambar 4.14 <i>Confusion Matrix</i> Bahasa Inggris	79
Gambar 4.15 <i>Confusion Matrix</i> Bahasa Spanyol	80
Gambar 4.16 <i>Confusion Matrix</i> Bahasa Prancis	80
Gambar 4.17 <i>Confusion Matrix</i> Bahasa Rusia	81
Gambar 4.18 <i>Confusion Matrix</i> Bahasa Turki	81
Gambar 4.19 <i>Confusion Matrix</i> Bahasa China.....	82

INTISARI

Indonesia, sebagai negara demokrasi rutin mengadakan pemilihan umum (pemilu) setiap lima tahun untuk memilih anggota legislatif dan presiden. Pada tahun 2024, pemilu kembali diadakan untuk memilih presiden dan wakil presiden, anggota DPR RI, DPD, DPRD Provinsi, serta DPRD Kabupaten/Kota. Topik pemilu menjadi perbincangan hangat, sehingga banyak portal berita memuat berita terkait pemilu. Analisis sentimen multibahasa dari berita *online* mengenai pemilu 2024 dimulai dengan *scraping* data dari berita di *detik.com* menggunakan kata kunci “pemilu 2024”. Data yang terkumpul diterjemahkan menggunakan *library googletrans*, kemudian dilabeli dengan metode *zero shot* sebagai sentimen negatif atau positif. Data yang tidak seimbang antara jumlah sentimen negatif dan positif disesuaikan menggunakan SMOTE (*Synthetic Minority Over-sampling Technique*). Proses klasifikasi dilakukan dengan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier*, sementara evaluasi menggunakan *Confusion matrix* untuk membandingkan performa kedua algoritma. Hasil analisis sentimen memberikan gambaran tentang pelaksanaan pemilu 2024. Penelitian ini menggunakan *dataset* sebanyak 7.181 berita dari *detik.com* yang diunggah antara 22 Oktober 2023 hingga 22 Februari 2024 dengan kata kunci “pemilu 2024”. Dari klasifikasi yang dilakukan, akurasi tertinggi dengan algoritma *Support Vector Machine* (SVM) pada *dataset* bahasa Inggris mencapai 78%.

Kata kunci: Pemilu 2024, Berita Online, Support Vector Machine, Multibahasa, Zero shot

ABSTRACT

Indonesia, as a democratic country, routinely holds general elections (pemilu) every five years to elect legislative members and the president. In 2024, the election will be held again to elect the president and vice president, members of the DPR RI, DPD, DPRD Province, and DPRD Regency/City. The topic of elections is a hot topic of discussion, so many news portals publish news related to elections. Multilingual sentiment analysis of online news about the 2024 election begins with scraping data from news on detik.com using the keyword “pemilu 2024”. The collected data is translated using the googletrans library, then labeled with the zero shot method as negative or positive sentiment. Data that is unbalanced between the amount of negative and positive sentiment is adjusted using SMOTE (Synthetic Minority Over-sampling Technique). The classification process is carried out using the Support Vector Machine (SVM) and Naïve Bayes Classifier algorithms, while the evaluation uses a Confusion matrix to compare the performance of the two algorithms. The results of the sentiment analysis provide an overview of the implementation of the 2024 election. This study used a dataset of 7,181 news articles from detik.com uploaded between October 22, 2023, and February 22, 2024, with the keyword “pemilu 2024”. From the classification carried out, the highest accuracy with the Support Vector Machine (SVM) algorithm on the English language dataset reached 78%.

Keywords: 2024 Elections, Online News, Support Vector Machine, Multilingual, Zero shot