

**ANALISIS DAN IMPLEMENTASI SISTEM SENTIMEN
TERHADAP ELEKTABILITAS CALON PRESIDEN
INDONESIA 2024 MENGGUNAKAN METODE SUPPORT
VECTOR MACHINE**

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data**



Disusun Oleh:

JASMINE AVRILE KANIASARI HARAHAP

20083010012

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2024**

LEMBAR PENGESAHAN

**ANALISIS DAN IMPLEMENTASI SISTEM SENTIMEN TERHADAP
ELEKTABILITAS CALON PRESIDEN INDONESIA 2024
MENGUNAKAN METODE SUPPORT VECTOR MACHINE**

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data
pada : Senin, 15 Juli 2024

**Program Studi S-1 Sains Data
Fakultas Ilmu Komputer**

**Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya**

Oleh:

JASMINE AVRILE KANIASARI HARAHAP

NPM. 20083010012

Disetujui oleh Tim Penguji Skripsi:

Penguji 1

Penguji 2

Dr.Eng.Ir. Dwi Arman Prasetya., ST., MT., IPU.
NIP. 198012052005011002

Aviolla Terza Damaliana, S.Si., M.Stat.
NIP. 199408022022032015

Pembimbing 1

Pembimbing 2

**Wahyu Syaifullah Jauharis Saputra., S.Kom.,
M.Kom**
NIP. 198608252021211003

Dr. Ir. Mohammad Idhom, SP., S.Kom., M.T
NIP. 198303102021211006

Mengetahui,

Fakultas Ilmu Komputer
Dekan,

Program Studi Sains Data
Fakultas Ilmu Komputer
Koordinator,

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 196811261994032001

Dr.Eng.Ir.Dwi Arman Prasetya.,ST.,MT.,IPU.
NIP. 198012052005011002

Surabaya, Juli, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Jasmine Avrile Kaniasari Harahap
NPM : 20083010012
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

**ANALISIS DAN IMPLEMENTASI SISTEM SENTIMEN TERHADAP
ELEKTABILITAS CALON PRESIDEN INDONESIA 2024
MENGUNAKAN METODE SUPPORT VECTOR MACHINE**

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 05 Juli 2024

Hormat Saya



Jasmine Avrile Kaniasari Harahap

NPM. 20083010012

ABSTRAK

ANALISIS DAN IMPLEMENTASI SISTEM SENTIMEN TERHADAP ELEKTABILITAS CALON PRESIDEN INDONESIA 2024 MENGUNAKAN METODE SUPPORT VECTOR MACHINE

Nama Mahasiswa / NPM : Jasmine Avrile Kaniasari Harahap / 20083010012
Program Studi : Sains Data, FASILKOM, UPN Veteran Jatim
Dosen Pembimbing 1 : Wahyu Syaifullah Jauharis Saputra., S.Kom., M.Kom
Dosen Pembimbing 2 : Dr. Ir. Mohammad Idhom, SP., S.Kom., M.T

Abstrak

Indonesia adalah negara yang menerapkan demokrasi dalam memilih calon presiden melalui proses pemilu. Dalam menanggapi calon-calon presiden masyarakat memiliki pandangan tersendiri mengenai calon presiden yang didukung. Pada era digital ini, menjadikan media sosial sebagai wadah masyarakat untuk menyampaikan opininya. Opini dari masyarakat dapat berupa opini positif maupun negatif, penggiringan opini publik, serta ujaran kebencian serta beragam komentar lainnya yang dapat menimbulkan permusuhan, penghinaan, perdebatan, dan perselisihan. Opini dari masyarakat dinilai penting karena dapat mempengaruhi pilihan masyarakat dan menentukan hasil akhir dari pemilu 2024. Dalam menanggapi permasalahan tersebut, diperlukan sebuah solusi untuk mengetahui seberapa banyak opini masyarakat terhadap calon presiden Indonesia cenderung ke opini positif atau negatif. Sehingga, dibuatlah sebuah sistem yang dapat melakukan klasifikasi sentimen dari opini yang disampaikan oleh masyarakat, dengan tujuan untuk mengklasifikasikan komentar masyarakat dari media sosial twitter yang berkaitan dengan pemilu 2024 ke dalam sentimen positif, dan negatif serta mengimplementasikannya ke dalam sebuah sistem berbasis *website*. Dalam penelitian ini, pemodelan data menggunakan metode *Support Vector Machine* (SVM) yang akan dievaluasi menggunakan *confusion matrix* sehingga didapatkan hasil nilai akurasi rata-rata terbaik sebesar 93.78%, nilai presisi sebesar 93.67%, nilai *recall* sebesar 93.78%, dan nilai *f1-score* sebesar 93.72% dengan kernel *rbf* dan nilai parameter *cost* (C) = 10 dan parameter *gamma* (γ) = 1 dengan pembagian proporsi data *train* dan data *test* 70:30.

Kata kunci: *Analisis Sentimen, Pemilu, Support Vector Machine.*

ABSTRACT

ANALYSIS AND IMPLEMENTATION OF A SENTIMENT SYSTEM ON THE ELECTABILITY OF INDONESIAN PRESIDENTIAL CANDIDATES 2024 USING THE SUPPORT VECTOR MACHINE METHOD

Student Name / NPM : Jasmine Avrile Kaniasari Harahap / 20083010012
Study Program : Sains Data, FASILKOM, UPN Veteran Jatim
Advisor 1 : Wahyu Syaifullah Jauharis Saputra., S.Kom., M.Kom
Advisor 2 : Dr. Ir. Mohammad Idhom, SP., S.Kom., M.T

Abstract

Indonesia is a country that implements democracy in choosing presidential candidates through the election process. In responding to presidential candidates, people have their own views about the presidential candidates they support. In this digital era, social media has become a place for people to express their opinions. These opinions can be in the form of positive or negative opinions, public opinion, and hate speech and various other comments that can cause hostility, humiliation, debate and dispute. It is important to consider public opinion, as it can influence people's choices and determine the final results of the 2024 elections. In response to these issues, it is necessary to identify the extent to which public opinion on Indonesian presidential candidates is positive or negative. A system was thus devised to classify sentiments expressed by the public in their opinions, with the objective of categorising public comments on Twitter related to the 2024 election into positive and negative sentiments and implementing them into a web-based system. In this study, the efficacy of data modelling using the Support Vector Machine (SVM) method will be evaluated using a confusion matrix, with the objective of achieving the optimal average accuracy value of 93.78%, precision value of 93.67%, and recall value of 93.78%. The results yielded a precision value of 93.67%, recall value of 93.78%, and an f1-score value of 93.72% with an rbf kernel and a cost parameter value (C) of 10 and a gamma parameter (γ) of 1, with a proportion of 70% for the training data and 30% for the test data.

Keywords: *Sentiment Analysis, Election, Support Vector Machine*

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat dan karunian-Nya penulis dapat menyelesaikan skripsi ini dengan judul “Analisis dan Implementasi Sistem Sentimen Terhadap Elektabilitas Calon Presiden Indonesia 2024 Menggunakan Metode Support Vector Machine”. Penulisan skripsi ini ditujukan untuk memenuhi syarat untuk memperoleh gelar Sarjana Sains Data pada program studi S1 Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Veteran Jawa Timur. Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada

1. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Prof. Dr. Ir. Novirina Hendrasarie, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur
3. Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Wahyu Syaifullah Jauharis Saputra., S.Kom., M.Kom. dan Dr. Ir. Mohammad Idhom, SP., S.Kom., M.T sebagai Dosen Pembimbing 1 dan 2.
5. Kedua orang tua dan keluarga yang senantiasa memberikan dukungan dan doa.
6. Teman – teman yang sudah menjadi *support system* ketika sedang suntuk dalam penyusunan laporan ini.
7. Kepada diri sendiri (Penulis) yang mampu dan kuat dalam menyelesaikan skripsi sampai tahap sidang akhir skripsi dengan segala situasi dan kondisi diri yang tidak tentu setiap waktunya.

Serta pihak-pihak lain yang tidak dapat disebutkan satu-persatu. Semoga laporan ini dapat dipergunakan dengan sebaik-baiknya

Surabaya, Juli 2024

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Dasar Teori.....	7
2.1.1 Analisis Sentimen	7
2.1.2 Twitter	7
2.1.3 <i>Lexicon Based</i>	8
2.1.4 <i>Text Preprocessing</i>	9
2.1.5 TF – IDF (<i>Term Frequency – Inverse Document Frequency</i>).....	10
2.1.6 <i>Chi-Square</i>	12
2.1.7 Support Vector Machine (SVM).....	13
2.1.8 <i>Confusion Matrix</i>	16
2.1.9 <i>Word Cloud</i>	18
2.2. Penelitian Terdahulu	18
BAB III METODOLOGI PENELITIAN.....	23
3.1. Variabel Penelitian dan Sumber Data	23

3.2.	Langkah Analisis.....	25
3.3.	Diagram Alir Penelitian	33
3.4.	Jadwal Penelitian.....	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Pengumpulan Data (<i>Crawling Data</i>).....	34
4.2	<i>Text Preprocessing</i>	37
4.2.1	<i>Cleaning Text</i>	37
4.2.2	<i>Case Folding</i>	38
4.2.3	<i>Tokenization</i>	39
4.2.4	Normalisasi	40
4.2.5	<i>Stopwords Removal</i>	41
4.2.6	<i>Stemming</i>	42
4.3	Labelisasi Data	42
4.4	TF-IDF	46
4.5	Fitur Seleksi <i>Chi-Square</i>	49
4.6	Proses Klasifikasi	51
4.6.1	Pembagian Data	51
4.6.2	Pemodelan Klasifikasi Support Vector Machine	53
4.7	Evaluasi	56
4.7.1	Pengujian Terhadap Data Sentimen Anies.....	56
4.7.2	Pengujian Terhadap Data Sentimen Prabowo.....	58
4.7.3	Pengujian Terhadap Data Sentimen Ganjar	60
4.8	Visualisasi <i>Word Cloud</i>	61
4.9	<i>Deployment</i>	65
BAB V PENUTUP.....		68
5.1	Kesimpulan	68
5.2	Saran.....	69
DAFTAR PUSTAKA		70
LAMPIRAN		74
BIODATA PENULIS		83

DAFTAR GAMBAR

Gambar 2. 1. <i>Hyperplane</i> SVM (Rizal dkk., 2019)	14
Gambar 2. 2. Kernel Linear (Praghakusma & Charibaldi, 2021)	15
Gambar 2. 3. Kernel <i>Polynomial</i> (Praghakusma & Charibaldi, 2021)	16
Gambar 3. 1. <i>Text Preprocessing</i>	25
Gambar 3. 2. Proses TF-IDF	29
Gambar 3. 3. Diagram Alir Penelitian	33
Gambar 4. 1. File CSV <i>dataset</i>	37
Gambar 4. 2. Hasil Proses Labelisasi Data	44
Gambar 4. 3. Hasil Sentimen Dataset Anies	44
Gambar 4. 4. Hasil Sentimen Dataset Prabowo	45
Gambar 4. 5. Hasil Sentimen Dataset Ganjar	45
Gambar 4. 6. <i>Confusion Matrix</i> Data Anies.....	57
Gambar 4. 7. <i>Confusion Matrix</i> Data Prabowo.....	59
Gambar 4. 8. <i>Confusion Matrix</i> Data Ganjar	60
Gambar 4. 9. <i>Wordcloud</i> Data Anies Sebelum <i>Preprocessing</i>	62
Gambar 4. 10. <i>Wordcloud</i> Data Anies Setelah <i>Preprocessing</i>	62
Gambar 4. 11. <i>Wordcloud</i> Data Prabowo Sebelum <i>Preprocessing</i>	63
Gambar 4. 12. <i>Wordcloud</i> Data Prabowo Setelah <i>Preprocessing</i>	64
Gambar 4. 13. <i>Wordcloud</i> Data Ganjar Sebelum <i>Preprocessing</i>	64
Gambar 4. 14. <i>Wordcloud</i> Data Ganjar Setelah <i>Preprocessing</i>	65
Gambar 4. 15. <i>Syntax Export Model</i>	66
Gambar 4. 16. Tampilan Halaman <i>Website</i>	66
Gambar 4. 17. Hasil Pengujian Aplikasi (1)	67
Gambar 4. 18. Hasil Pengujian Aplikasi (2)	67

DAFTAR TABEL

Tabel 2. 1. <i>Confusion Matrix</i>	17
Tabel 2. 2. Studi Literatur	19
Tabel 3. 1 Tabel Struktur Data Penelitian	23
Tabel 3. 2 Tabel Lanjutan Kolom Struktural Data Penelitian.....	24
Tabel 3. 3. Contoh <i>Dataset</i> Untuk <i>Cleaning Text</i>	26
Tabel 3. 4 Contoh <i>Dataset</i> untuk <i>Case Folding</i>	26
Tabel 3. 5 Contoh <i>Dataset Tokenization</i>	27
Tabel 3. 6 Contoh <i>Dataset</i> Normalisasi	27
Tabel 3. 7 Contoh <i>Dataset</i> untuk <i>Stopword Removal</i>	27
Tabel 3. 8 Contoh Data untuk <i>Stemming</i>	28
Tabel 3. 9 Contoh Labelisasi Data menggunakan Inset.....	28
Tabel 3. 10 Contoh <i>Dataset</i> Penelitian	30
Tabel 3. 11 Kemunculan dan Ketidakhadiran <i>Term</i> pada Tiap Kelas.....	30
Tabel 3. 12 Jadwal Penelitian	33
Tabel 4. 1. Data Penelitian	36
Tabel 4. 2. Proses <i>Cleaning Text</i>	38
Tabel 4. 3. Proses <i>Case Folding</i>	38
Tabel 4. 4. Proses <i>Tokenization</i>	39
Tabel 4. 5. Proses Normalisasi.....	40
Tabel 4. 6. Proses <i>Stopword Removal</i>	41
Tabel 4. 7. Proses <i>Stemming</i>	42
Tabel 4. 8. Data Sampel Perhitungan TF IDF.....	46
Tabel 4. 9. Perhitungan TF IDF pada Sampel Data	47
Tabel 4. 10. Hasil Pembobotan TF-IDF.....	48
Tabel 4. 11. Hasil Statistik Uji <i>Chi-Square</i>	49
Tabel 4. 12. Hasil Seleksi Fitur Data Anies	50
Tabel 4. 13. Hasil Seleksi Fitur Data Prabowo	50
Tabel 4. 14. Hasil Seleksi Fitur Data Ganjar	51
Tabel 4. 15. Pembagian Data 70:30 Data Anies	52
Tabel 4. 16. Pembagian Data 80:20 Data Anies	52
Tabel 4. 17. Pembagian Data 90:10 Data Anies	52

Tabel 4. 18. Pembagian Data 70:30 Data Prabowo	52
Tabel 4. 19. Pembagian Data 80:20 Data Prabowo	52
Tabel 4. 20. Pembagian Data 90:10 Data Prabowo	52
Tabel 4. 21. Pembagian Data 70:30 Data Ganjar.....	53
Tabel 4. 22. Pembagian Data 80:20 Data Ganjar.....	53
Tabel 4. 23. Pembagian Data 90:10 Data Ganjar.....	53
Tabel 4. 24. Hasil Parameter Terbaik dan Seleksi Fitur Data Sentimen Anies	54
Tabel 4. 25. Hasil Parameter Terbaik dan Seleksi Fitur Data Sentimen Prabowo	55
Tabel 4. 26. Hasil Parameter Terbaik dan Seleksi Fitur Data Sentimen Ganjar ...	55
Tabel 4. 27. <i>Classification Report</i> Data Anies.....	57
Tabel 4. 28. <i>Classification Report</i> Data Prabowo.....	58
Tabel 4. 29. <i>Classification Report</i> Data Ganjar	60

DAFTAR LAMPIRAN

Lampiran 1. Hasil Uji Plagiasi	74
Lampiran 2. Data Penelitian.....	75
Lampiran 3. <i>Source Code</i> yang digunakan untuk analisis	76