

Penerapan SVM dengan Seleksi Fitur Mutual Information untuk Memprediksi Sentimen PEMILU 2024

I Gusti Bgs Darmika Putra^{a1}, Cokorda Pramatha^{a2},
Anak Agung Istri Ngurah Eka Karyawati^{a3}, Made Agung Raharja^{a4}

^aProgram Studi Informatika, Universitas Udayana

^bCenter for Interdisciplinary Research on the Humanities and Social Sciences, Universitas Udayana
Bali, Indonesia

¹gstbgsdarmika@gmail.com

²cokorda@unud.ac.id

³eka.karyawati@unud.ac.id

⁴made.agung@unud.ac.id

Abstract

A wealth of information on the 2024 Indonesian Election floods Twitter, from campaign schedules to candidate profiles and the latest survey results indicating candidate popularity. This information overload poses challenges in discerning comments' sentiment. Manual classification is feasible but time-consuming. Hence, this study aims to streamline data analysis for the 2024 Election. It employs a dataset of 1000 entries categorized as positive or negative. Support Vector Machine (SVM) with Mutual Information feature selection is utilized for classification. Results reveal that Mutual Information feature selection enhances SVM performance. Without it, SVM achieves 88% accuracy and 87.9% f-measure using the rbf kernel ($C=1$, $\gamma=2$), computed in about 0.07 seconds. With feature selection, SVM's accuracy improves to 90%, and f-measure to 89.9% with 60% features, using rbf kernel ($C=10$, $\gamma=0.5$), reducing computation time to 0.02 seconds, optimizing both performance and efficiency. The website system scored 88.63 in usability, higher than the global average of 68, based on a SUS questionnaire with 10 questions and 20 respondents. This indicates excellent performance and user satisfaction, as evaluated from the web system.

Keywords: Sentiment Analysis, Classification, Support Vector Machine, Mutual Information, Indonesian Election 2024

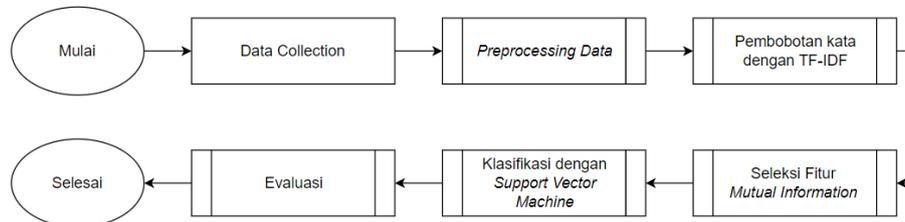
1. Pendahuluan

Pemilihan umum merupakan suatu proses dalam sistem demokrasi yang melibatkan partisipasi aktif masyarakat dalam menentukan wakil-wakil para perwakilannya di lembaga-lembaga pemerintahan [1]. Banyak informasi mengenai Pemilu 2024 tersebar di *Twitter*, mulai dari jadwal kampanye, profil pasangan calon, hingga hasil survei terbaru yang menunjukkan seberapa populer masing-masing kandidat [2]. Hal ini menyulitkan untuk menentukan apakah suatu komentar bersifat positif atau negatif, meskipun dapat dilakukan secara manual, namun memerlukan waktu dan usaha yang lebih. Analisis sentimen hadir sebagai solusi yang mempermudah dalam memahami reaksi orang terhadap suatu topik, dan dengan demikian dapat mendukung pengambilan keputusan yang lebih tepat. Dalam analisis sentimen, data biasanya dikelompokkan menjadi positif, negatif, atau netral untuk memudahkan pemahaman mengenai opini atau tanggapan yang terkandung dalam teks [3]. Sebagai contoh, dalam penelitian mengenai analisis sentimen tweet terkait program kartu prakerja [4]. Hasil eksperimen tersebut menunjukkan bahwa SVM dapat digunakan secara efektif untuk mengklasifikasikan tweet dengan Tingkat akurasi mencapai 98.67%, *precision* 98%, *recall* 99%, dan *f-measure* 98%. Terdapat juga penelitian yang melakukan analisis sentimen terhadap ulasan produk berdasarkan aspek-aspeknya, menggunakan pemilihan fitur *Mutual Information* (MI) dengan pendekatan *K-Nearest Neighbor* (KNN) oleh [5]. Berdasarkan hasil eksperimen, ditemukan bahwa penggunaan MI memberikan pengaruh yang relatif baik terhadap peningkatan akurasi pada aspek harga 39,77%, aspek kemasan 7,31%, dan 9,18% pada aspek aroma. Berdasarkan uraian dan referensi penelitian sebelumnya, penulis akan menganalisis sentimen pada data *tweet* terkait pemilu 2024 menggunakan metode SVM dengan seleksi fitur *Mutual Information* (MI). Penulis memilih kasus pemilu dalam

penelitian ini karena pemilu merupakan proses demokrasi yang penting, mencerminkan partisipasi masyarakat dalam menentukan pemimpin dan kebijakan publik. Dengan menggunakan kombinasi ini, diharapkan penelitian ini bisa menjadi panduan bagi perbandingan pilihan metode pendekatan dalam menganalisis sentimen terhadap opini di *Twitter*.

2. Metode Penelitian

Pada bagian metodologi penelitian, penulis menerapkan serangkaian pendekatan atau metode untuk melakukan analisis sentimen. Langkah-langkah yang diambil oleh penulis dalam penyusunan penelitian ini mencakup pencarian dan pengumpulan data, preprocessing, pelabelan data, pembobotan data dan klasifikasi metode *Support Vector Machine* (SVM) dengan seleksi fitur *Mutual Information* (MI). Alur penelitian dapat dilihat pada Gambar 1 berikut:



Gambar 1. Diagram Alir Penelitian

2.1. Pengumpulan Data

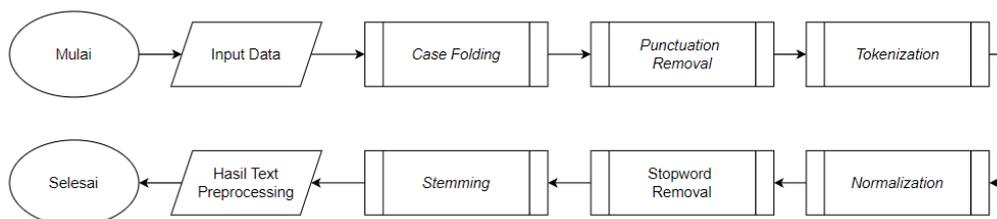
Penelitian ini menggunakan data sekunder dari media sosial *Twitter* berbahasa Indonesia dengan kata kunci "pemilu 2024". Total data yang digunakan adalah 1000 *tweet* yang terbagi menjadi dua kelas, yaitu positif dan negatif, dengan masing-masing 500 *tweet*. Pengumpulan data dilakukan mulai bulan Mei hingga Desember 2024. Proses pelabelan data dilakukan secara manual dengan melibatkan ahli bahasa Indonesia dari Fakultas Ilmu Budaya Universitas Udayana. Contoh data untuk setiap kelas dapat dilihat pada Tabel 1 berikut:

Tabel 1. Contoh Data Tweet

Tweet	Label
Pemilu 2024 adalah kesempatan bagi kita semua untuk membuktikan bahwa demokrasi Indonesia bisa menjadi contoh bagi dunia @hidupdamaiiiii Ganjar Mahfud Menang https://t.co/QOKmnZBMQX	Positif
#TolakKetuaKPUCurang ... tolak ketua KPU CURANG...semoga cepat mati sebelum pemilu 2024.	Negatif

2.2. Text Preprocessing

Text preprocessing adalah tahap awal dalam proses klasifikasi teks yang bertujuan untuk mengubah data teks asli yang tidak terstruktur menjadi data yang terstruktur, serta memastikan data yang berkualitas atau inputan yang baik agar dapat dilanjutkan ke proses selanjutnya dengan efisien [6]. Pada penelitian ini, dilakukan beberapa tahap *text preprocessing* yang dijelaskan pada Gambar 2:



Gambar 2. Tahapan *Text Preprocessing*

Tahapan pertama dalam *text preprocessing* adalah *case folding*, yaitu mengubah semua karakter menjadi huruf kecil. Dilanjutkan dengan *punctuation removal* untuk penghapusan tanda baca yang diterapkan dalam suatu teks yang bertujuan untuk membedakan antara kalimat dan bagian penyusunnya, serta untuk menjelaskan maknanya dengan lebih jelas. Kemudian, teks dibagi menjadi token dengan proses *tokenization*, diikuti oleh *normalization* untuk mengonversi variasi bentuk kata menjadi bentuk standar. Setelah itu, langkah selanjutnya adalah *stopword removal* untuk

menghilangkan kata-kata umum yang kurang relevan. Terakhir, tahap *stemming* dilakukan proses mengubah kata menjadi bentuk dasarnya dengan menghapus akhiran kata. Ini membantu mengurangi variasi dasar antara kata yang serupa [7].

2.3. Term Frequency – Inverse Document Frequency

Pembobotan kata merupakan suatu proses yang bertujuan untuk mengurangi dampak dari kata-kata yang umum atau tidak relevan dalam sebuah dokumen, sementara pada saat yang sama memberikan bobot yang lebih besar untuk kata-kata yang dianggap penting atau memiliki makna yang lebih spesifik. Salah satu algoritma yang umum digunakan adalah TF-IDF [8]. Adapun langkah perhitungannya sebagai berikut:

a. Menghitung *term frequency*

$$TF_{(t,d)} = \frac{\text{jumlah kata } t \text{ dalam dokumen } d}{\text{jumlah kata total dalam dokumen } d} \quad (1)$$

b. Menghitung *inverse document frequency*

$$IDF_{(t)} = \log \frac{N}{DF_{(t)}} \quad (2)$$

c. Menghitung bobot TF-IDF

$$TF - IDF_{(t,d)} = TF(t,d) \times IDF(t) \quad (3)$$

Keterangan:

d = dokumen yang sedang dihitung.

t = Kata yang diukur frekuensinya.

N = Jumlah total dokumen.

IDF(t) = frekuensi dokumen dari term t.

TF(t,d) = Nilai *Term Frequency* kata (t) dalam dokumen (d).

IDF(t) = Nilai IDF kata (t) dalam kumpulan dokumen.

2.4. Mutual Information

Mutual Information dalam analisis sentimen membantu dalam memahami keterkaitan antara fitur-fitur yang muncul dalam teks dengan sentimen yang diungkapkan [9]. Hal ini memungkinkan untuk membangun model analisis sentimen yang lebih akurat dan dapat memberikan wawasan yang lebih baik tentang sentimen yang terkandung dalam teks. Untuk mencari nilai *Mutual Information* pada setiap kelas kata dalam kategori, menggunakan rumus berikut:

$$I(U, C) = \sum_{et \in \{1,0\}, ec \in \{1,0\}} P(U = et, C = ec) \log_2 \frac{P(U = et, C = ec)}{P(U = et) P(C = ec)} \quad (4)$$

Persamaan (4) dapat dijabarkan kembali menjadi persamaan (5). Berikut merupakan persamaan dari *Mutual Information* yang dijabarkan kembali.

$$I(U, C) = \frac{N_{11}}{N} \log_2 \frac{N \cdot N_{11}}{N_1 \cdot N_1} + \frac{N_{01}}{N} \log_2 \frac{N \cdot N_{01}}{N_0 \cdot N_1} + \frac{N_{10}}{N} \log_2 \frac{N \cdot N_{10}}{N_1 \cdot N_0} + \frac{N_{00}}{N} \log_2 \frac{N \cdot N_{00}}{N_0 \cdot N_0} \quad (5)$$

Keterangan:

N = Total dokumen yang memiliki et dan ec.

N₁ = Total dokumen yang memiliki et.

N₀ = Total dokumen yang memiliki ec.

N₀ = Total dokumen yang memiliki tidak memiliki et.

N₀ = Total dokumen yang memiliki tidak memiliki et.

2.5. Support Vector Machine

Support Vector Machine (SVM) adalah algoritma klasifikasi data yang termasuk dalam kategori *supervised learning*. Dalam penggunaannya, SVM memerlukan data yang telah dilabeli sebelumnya sebagai masukan. SVM bertujuan memisahkan data ke dalam kategori berbeda dengan menemukan

garis pemisah optimal [10]. SVM menggunakan konsep kernel dalam dimensi tinggi untuk memecahkan masalah *non-linear*. Konsep ini memungkinkan SVM untuk menemukan hubungan yang kompleks antara fitur dalam data, sehingga meningkatkan kemampuan SVM dalam menangani berbagai jenis data yang tidak dapat dipisahkan secara *linear* [11]. Formulasi optimasi SVM dibagi menjadi klasifikasi linear dan non-linear. Berikut merupakan rumus SVM:

- a. Menghitung garis *hyperplane* menggunakan rumus persamaan

$$w_i \cdot x + b = 0 \quad (6)$$

- b. Jika data termasuk kelas positif menggunakan rumus persamaan

$$w_i \cdot x + b > 0 \quad (7)$$

- c. Jika data termasuk kelas negatif menggunakan rumus persamaan

$$w_i \cdot x + b < 0 \quad (8)$$

- d. Dengan adanya dua garis pemisah tersebut maka dapat menghasilkan persamaan

$$y_i((\vec{x}_i) \cdot \vec{w} + b - 1 \geq 0) \quad (9)$$

Keterangan:

w = Vektor bobot (*weight vector*) yang merupakan vektor normal terhadap garis *hyperplane*.

x = Vektor fitur *input* yang merupakan data yang ingin diklasifikasikan.

b = Bias atau pergeseran *hyperplane*.

y_i = Label kelas dari data latih.

\vec{x}_i = Vektor fitur *input* dari data latih.

\vec{w} = Vektor bobot yang digunakan untuk memisahkan dua kelas data.

2.6. System Usability Scale

SUS merupakan alat pengukuran standar yang digunakan untuk mengevaluasi tingkat kegunaan (*usability*) suatu sistem [12]. Salah satu keunggulan SUS adalah kemampuannya untuk memberikan hasil yang cukup signifikan dengan jumlah sampel yang relatif sedikit, sehingga dapat mengurangi biaya pengujian *usability* secara keseluruhan. Berikut merupakan rumus *System Usability Scale* (SUS):

$$\text{Skor SUS} = ((P1 - 1) + (5 - P2) + (P3 - 1) + (5 - P4) + (P5 - 1) + (5 - P6) + (P7 - 1) + (5 - P8) + (P9 - 1) + (5 - P10)) * 2.5 \quad (10)$$

2.7. Uji Validitas

Uji validitas adalah langkah untuk memeriksa seberapa baik alat yang digunakan dalam penelitian bisa mengukur variabel yang seharusnya diukur dengan benar dan tepat [13]. Dengan kata lain, validitas membantu memastikan bahwa alat tersebut benar-benar mengukur apa yang seharusnya diukur dengan akurat. Uji validitas melibatkan dua metode utama: *Expert Review* dan uji *Product-Moment Coefficient*. *Expert Review* adalah evaluasi oleh ahli dalam bidang terkait, yang diminta untuk menilai apakah instrumen yang telah dibuat sudah valid atau belum [13].

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{(n(\sum x_i^2) - (\sum x_i)^2)(n(\sum y_i^2) - (\sum y_i)^2)}} \quad (11)$$

Keterangan:

r = koefisien korelasi.

n = jumlah responden.

x_i = skor setiap item pada instrument.

y_i = skor setiap item pada kriteria.

Setelah menemukan koefisien korelasi, langkah selanjutnya untuk mengetahui validitas dapat melakukannya dengan membandingkan nilai t_{hitung} dengan nilai t_{tabel} . Nilai t_{tabel} didapatkan sesuai dengan tingkat signifikansi penelitian yang telah ditentukan sebelumnya, dalam kasus ini adalah 5% atau 0.05. Sebuah butir dianggap valid jika nilai t_{hitung} lebih besar dari nilai t_{tabel} . Sedangkan nilai t_{hitung} didapatkan dengan menggunakan rumus [13].

$$r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (12)$$

Keterangan:

r = koefisien korelasi

n = jumlah responden

t = nilai t_{hitung}

2.8. Uji Reliabilitas

Uji reliabilitas adalah proses menilai seberapa konsisten suatu metode atau instrumen pengukuran dalam menghasilkan hasil yang serupa dalam situasi yang berbeda [13]. Dalam konteks kuesioner, uji reliabilitas penting untuk memastikan konsistensi pertanyaan dalam mengukur variabel yang sama. Ini memungkinkan hasil kuesioner tetap dapat diandalkan dan konsisten dalam berbagai situasi, mendukung analisis yang akurat dalam penelitian. Nilai yang diperoleh dari uji *Cronbach's Alfa* biasanya berada dalam rentang antara lebih dari 0,7 dan kurang dari 0,9. Semakin tinggi nilai *Cronbach's Alfa*, semakin reliabel instrumen penelitian tersebut dianggap [13]. Pengujian reliabilitas menggunakan rumus *Cronbach's Alfa* dengan rumus:

$$r_i = \frac{k}{(k - 1)} \left(1 - \frac{\sum s_i^2}{s_t^2} \right) \quad (13)$$

Keterangan:

r_i = koefisien reliabilitas *Cronbach's Alfa*.

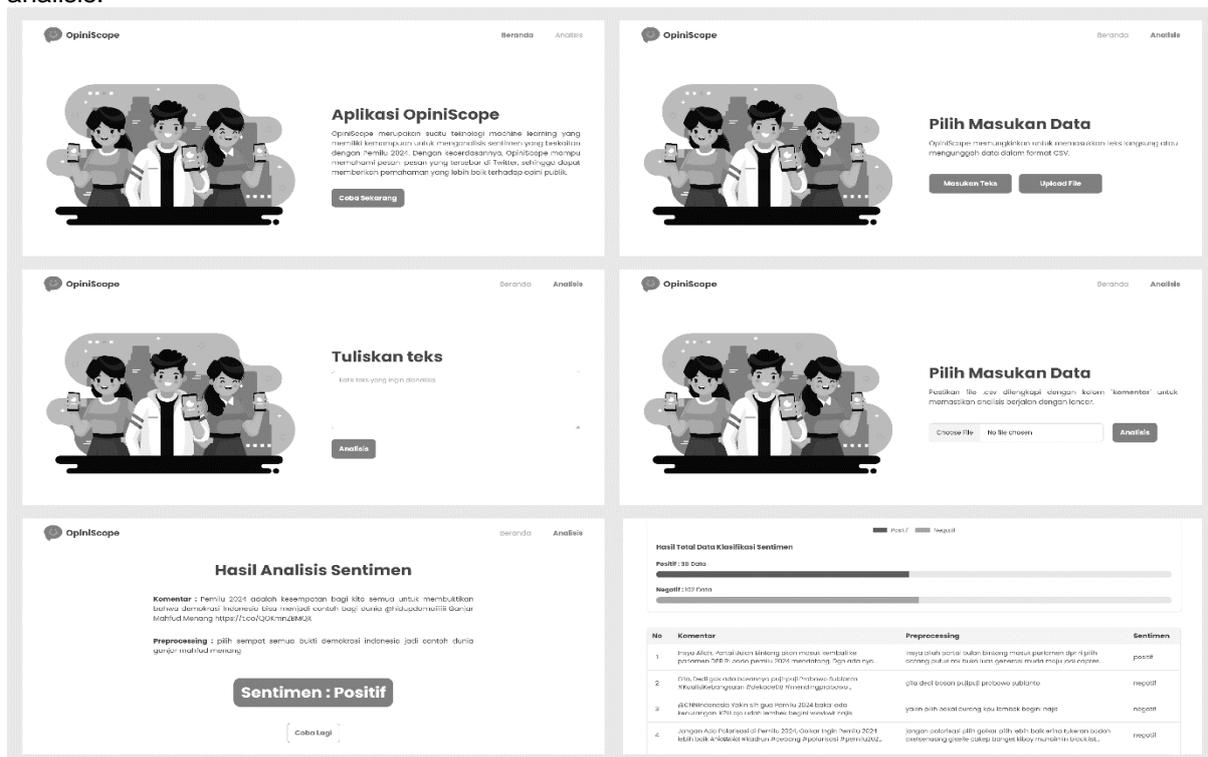
k = jumlah item soal.

s_i^2 = varians tiap item.

s_t^2 = varians total.

2.9. Rancangan Antarmuka

Rancangan antarmuka (*interface design*) merupakan proses perencanaan dan pengembangan tampilan program aplikasi yang dapat dilihat dan digunakan oleh pengguna [14]. Antarmuka sistem pada penelitian ini dibangun berbasis website menggunakan *Python* dengan *framework Flask* dan *Preline UI* sebagai *frontend*, yang menggunakan *Tailwind CSS component library*. Gambar 3 adalah hasil dari proses rancangan antarmuka sistem halaman beranda, halaman analisis, dan halaman hasil analisis.



Gambar 3. Rancangan Antarmuka

2.10. Skenario Pengujian dan Evaluasi

Penelitian ini menggunakan *K-Fold Cross-Validation* untuk pengujian model klasifikasi. Dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Kemudian, data latih dilakukan iterasi sebanyak k bagian atau folds, setiap bagian dibagi menjadi data latih dan data validasi secara bergantian. Tujuan pengujian ini adalah untuk menemukan parameter rasio seleksi fitur yang optimal dengan memperhatikan rata-rata nilai metrik seperti akurasi dan *f-measure* dari seluruh iterasi terhadap berbagai rasio seleksi fitur yang diujikan [15]. Selanjutnya, kinerja parameter rasio seleksi fitur terbaik akan dibandingkan saat diterapkan pada data uji, baik dengan menggunakan seleksi fitur maupun tanpa penggunaan seleksi fitur. Sedangkan untuk evaluasi sistem website menggunakan kuesioner SUS dengan skala *Likert*. Kuesioner ini dirancang untuk mendapatkan persepsi responden terhadap *usability* (kemudahan penggunaan) website. Kegunaan diukur dari seberapa percaya seseorang bahwa penggunaan sistem tersebut akan meningkatkan kinerja kerjanya. Kemudahan penggunaan diukur dari seberapa percaya seseorang bahwa menggunakan sistem tersebut akan mudah tanpa memerlukan usaha [16].

3. Hasil dan Pembahasan

3.1. Pembangunan Model Klasifikasi

Terdapat 2 skenario klasifikasi, yaitu dengan menggunakan seleksi fitur dan tanpa seleksi fitur. Terdapat 3 *tuning hyperparameter* yang digunakan yaitu: kernel, C , dan γ (γ). Adapun kernel yang diuji meliputi *linear*, *rbf*, dan *polynomial*. Rentang nilai yang dievaluasi untuk parameter C adalah 0.1, 1, dan 10, sementara untuk γ (γ) adalah 0.5, 1, dan 2. Pengujian dilakukan dengan menggunakan metode *K-fold validation*, di mana nilai *K-fold validation* yang digunakan adalah 5. Berikut adalah tabel perbandingan akurasi dan *f-measure* antara penggunaan metode seleksi fitur dan tanpa seleksi fitur dalam klasifikasi SVM. Hasil evaluasi klasifikasi tanpa seleksi fitur ditunjukkan pada Tabel 2 sedangkan hasil evaluasi dengan seleksi fitur ditunjukkan pada Tabel 3.

Tabel 2. Hasil Evaluasi Klasifikasi Tanpa Seleksi Fitur

Kernel	C	γ	Jumlah Fitur	Accuracy	F-measure	Waktu Komputasi
Linear	1	-	1235	0.865	0.864	0.04
Rbf	1	2	1235	0.880	0.879	0.07
Polynomial	10	0.5	1235	0.835	0.839	0.06

Tabel 3. Hasil Evaluasi Klasifikasi dengan Seleksi Fitur

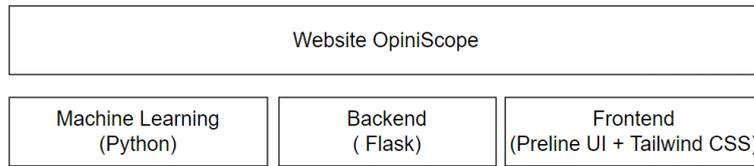
Kernel	Nilai Threshold	C	γ	Jumlah Fitur	Accuracy	F-measure	Waktu Komputasi
Linear	60%	10	-	741	0.890	0.889	0.02
Rbf	60%	10	0.5	741	0.900	0.899	0.02
Polynomial	80%	10	1	984	0.875	0.873	0.05

Pada hasil klasifikasi tanpa menggunakan seleksi fitur, ditemukan bahwa nilai tertinggi untuk akurasi dan *f-measure* tercapai saat menggunakan kernel *rbf* dengan parameter $C = 1$ dan $\gamma = 2$, dengan total fitur sebanyak 1235. Hasilnya menunjukkan akurasi sebesar 88% dan *f-measure* sebesar 87,9%, dengan waktu komputasi yang mencapai 0,07 detik. Sementara itu, dalam pengujian klasifikasi sentimen dengan menerapkan seleksi *Mutual Information* (MI), nilai akurasi dan *f-measure* tertinggi tercapai dengan menerapkan kernel *rbf* dengan seleksi fitur sebesar 60%, parameter $C = 10$, $\gamma = 0.5$ dengan jumlah fitur yang digunakan sebanyak 741 fitur yang menghasilkan akurasi sebesar 90% dan *f-measure* sebesar 89,9%, dengan waktu komputasi yang lebih efisien, yaitu 0,02 detik. Pengaruh signifikan dari seleksi fitur dalam klasifikasi menggunakan SVM terlihat jelas, di mana penerapan seleksi fitur mampu meningkatkan performa klasifikasi dengan mengurangi jumlah fitur yang digunakan. Ini tidak hanya mengoptimalkan waktu komputasi, tetapi juga dapat meningkatkan akurasi dan nilai *f-measure* secara keseluruhan.

3.2. Arsitektur Sistem

Arsitektur Sistem website yang dikembangkan mencakup penggunaan *Python* versi 3.12.1 sebagai dasar utamanya untuk membangun *machine learning* klasifikasi guna menganalisis data opini. *Flask* versi 3.0.2 digunakan di bagian *backend* aplikasi untuk menangani permintaan dan pengelolaan data di sisi server, sementara tampilan depannya diperkuat dengan *Preline UI* versi 2.0.0 yang

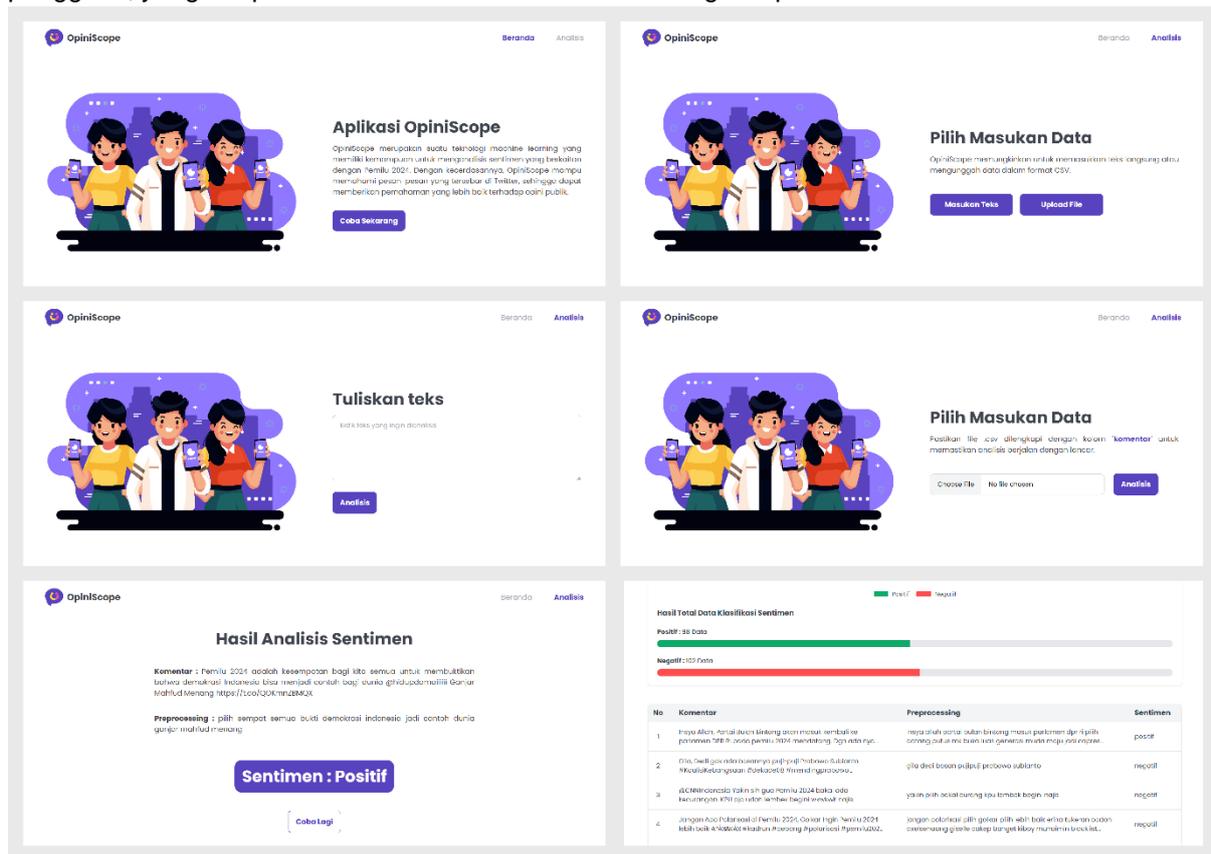
menggunakan *Tailwind CSS component library* untuk tampilan yang responsif dan menarik. Arsitektur sistem dapat dilihat pada Gambar 4 berikut:



Gambar 4. Arsitektur Sistem

3.3. Implementasi Antarmuka

Implementasi sistem ini akan dijelaskan sesuai dengan urutan tahapan yang telah ditetapkan sebelumnya. Sistem ini dirancang untuk memberikan fleksibilitas dalam hal jenis dan jumlah *input* komentar yang dapat digunakan serta memberikan hasil yang relevan dan mudah dimengerti bagi pengguna, yang direpresentasikan melalui Gambar 5 sebagai implementasi antarmuka sistem.



Gambar 5. Implementasi Antarmuka Sistem

3.4. Pengujian System Usability Scale (SUS)

Proses evaluasi website menggunakan kuesioner SUS dengan skala *Likert*. Kuesioner ini dirancang untuk mendapatkan persepsi responden terhadap *usability* website. Responden diminta untuk menilai sejumlah pernyataan yang diberikan dengan menggunakan skala yang ditentukan. Selanjutnya, hasil tanggapan responden akan diuji validitas dan reliabilitasnya. Langkah-langkah ini penting untuk memastikan bahwa evaluasi *usability* website didasarkan pada data yang valid dan dapat dipercaya. Berikut adalah sepuluh pernyataan yang digunakan untuk menguji sistem *OpiniScope* menggunakan metode SUS, yang tertera pada Tabel 4.

Tabel 4. Kuisisioner System Usability Scale

No	Pernyataan
P1	Saya akan sering mengakses website <i>OpiniScope</i> .
P2	Saya merasa website <i>OpiniScope</i> sulit untuk digunakan.

P3	Saya merasa website <i>OpiniScope</i> mudah digunakan.
P4	Saya butuh bantuan dari orang lain atau teknisi untuk menggunakan website <i>OpiniScope</i> .
P5	Saya merasa bahwa fitur-fitur situs website <i>OpiniScope</i> berjalan dengan baik.
P6	Saya merasa ada banyak hal yang tidak konsisten di website <i>OpiniScope</i> .
P7	Saya yakin orang lain cepat memahami cara menggunakan website <i>OpiniScope</i> .
P8	Saya merasakan kebingungan dalam menggunakan website <i>OpiniScope</i> .
P9	Saya merasa tidak mengalami kesulitan dalam menggunakan website <i>OpiniScope</i> .
P10	Saya merasa perlu membiasakan diri sebelum menggunakan website <i>OpiniScope</i> .

Evaluasi sistem pada penelitian ini melibatkan 20 responden sebagai partisipan. Di bawah ini adalah hasil evaluasi skor SUS yang diperoleh dari responden, yang tercantum pada Tabel 5:

Tabel 5. Hasil Perhitungan Skor SUS

Resp.	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Jumlah	Skor SUS
1	0	3	3	3	4	3	4	3	3	3	29	72,5
2	4	4	3	3	4	3	4	4	4	3	36	90
3	4	4	4	4	4	4	4	4	4	4	40	100
4	4	4	4	4	4	4	4	4	4	4	40	100
5	3	4	4	4	4	4	4	4	4	3	38	95
6	4	4	3	4	4	3	4	4	3	4	37	92,5
7	3	4	4	4	4	4	4	4	4	4	39	97,5
8	3	3	3	3	4	3	3	3	3	4	32	80
9	2	4	4	4	4	4	3	4	4	3	36	90
10	3	4	4	3	3	4	4	4	3	3	35	87,5
11	3	3	4	4	4	4	4	4	4	4	38	95
12	3	4	4	3	4	3	4	4	4	3	36	90
13	2	3	4	4	3	3	3	4	3	3	32	80
14	4	4	4	4	4	3	3	4	4	3	37	92,5
15	3	4	3	3	3	3	3	3	3	2	30	75
16	2	4	4	3	3	4	3	3	3	3	32	80
17	4	4	4	4	4	4	4	4	4	4	40	100
18	3	3	3	3	3	2	3	3	3	0	26	65
19	4	4	4	4	2	2	4	4	4	4	36	90
20	4	4	4	4	4	4	4	4	4	4	40	100
Skor Rata-Rata SUS											88,63	

Berdasarkan hasil perhitungan, website *OpiniScope* memperoleh nilai SUS sebesar 88,63 yang menunjukkan bahwa pengguna menganggap sistem ini dapat diterima dalam penggunaannya. Tingkat kegunaan sistem tersebut dinilai sebagai "Acceptable" berdasarkan *acceptability ranges*. Pada *grade scale*, website *OpiniScope* mendapatkan peringkat B, menunjukkan bahwa secara keseluruhan, tingkat kegunaan sistem tersebut berada pada tingkat yang baik. Peringkat B menunjukkan bahwa secara keseluruhan, website memiliki kualitas yang baik dan mudah digunakan. Sedangkan untuk *adjective rating* termasuk ke kelompok "Excellent" menunjukkan bahwa website memiliki tingkat kualitas yang sangat baik dalam hal usability dan pengalaman pengguna.

3.5. Hasil Uji Validitas

Dalam penelitian ini, validitas diukur menggunakan koefisien korelasi *Pearson* (2 tail) terhadap skor total dari 10 item pernyataan dalam kuesioner SUS. Untuk menentukan validitas, penulis melakukan perbandingan antara nilai t_{hitung} dengan nilai t_{tabel} pada tingkat signifikansi 0,05. Dalam kasus ini,

t_{tabel} memiliki nilai 0,444 karena jumlah responden adalah 20. Setelah melakukan pengukuran, didapatkan bahwa nilai t_{hitung} lebih besar dari t_{tabel} . Hal ini menunjukkan bahwa kuesioner memiliki validitas yang baik, karena hubungan antara setiap item dalam kuesioner dengan skor totalnya signifikan secara statistik. Berdasarkan hasil uji validitas, pernyataan dalam kuesioner dianggap valid karena nilai t_{hitung} yang dihasilkan lebih besar daripada t_{tabel} yang ditentukan. Rincian perhitungan lebih lanjut dapat ditemukan pada Tabel 6, yang memberikan penjelasan tentang hasil uji validitas data kuesioner.

Tabel 6. Hasil Uji Validitas Data Kuesioner

	t_{hitung}	t_{tabel}	Keterangan
P1	0.613	0.444	Valid
P2	0.593	0.444	Valid
P3	0.628	0.444	Valid
P4	0.714	0.444	Valid
P5	0.469	0.444	Valid
P6	0.619	0.444	Valid
P7	0.615	0.444	Valid
P8	0.827	0.444	Valid
P9	0.792	0.444	Valid
P10	0.764	0.444	Valid

3.6. Hasil Uji Reliabilitas

Hasil uji reliabilitas yang terdapat pada Tabel 7, menunjukkan bahwa kuesioner SUS memiliki tingkat reliabilitas yang cukup tinggi yaitu 0.830. Standar umum menyatakan bahwa nilai *Cronbach's Alpha* yang berada di antara 0.7 dan 0.9 menunjukkan tingkat reliabilitas yang baik. Dengan nilai 0.830 yang berada dalam rentang ini, dapat disimpulkan bahwa kuesioner tersebut dianggap reliabel. Ini berarti pertanyaan pada kuesioner dapat diandalkan untuk mengukur apa yang dimaksudkan dengan konsistensi yang baik.

Tabel 7. Hasil Uji Reliabilitas Data Kuesioner

Nilai Cronbach's Alfa	Jumlah Item	Keterangan
0.830	10	Reliabel

4. Kesimpulan

Hasil Penelitian ini, menunjukkan bahwa seleksi fitur *Mutual Information* (MI) dapat meningkatkan kinerja metode SVM dibandingkan tanpa seleksi fitur. Pada pengujian tanpa seleksi fitur, kinerja SVM mencapai akurasi 88% dan *f-measure* 87,9% dengan kernel *rbf* menggunakan nilai $C = 1$ dan nilai $\gamma = 2$ dengan waktu komputasi sekitar 0,07 detik. Saat seleksi fitur diterapkan, terjadi peningkatan kinerja SVM yang cukup signifikan. Akurasi meningkat menjadi 90%, sedangkan *f-measure* mencapai 89,9% dengan pemanfaatan 60% fitur, kernel *rbf*, dan parameter $C = 10$ serta $\gamma = 0,5$. Selain mengoptimalkan kinerja, penerapan seleksi fitur juga memberikan dampak positif pada efisiensi waktu komputasi, yang turun drastis menjadi hanya 0,02 detik.

References

- [1] Imam Fahrur Rozi, Yushintia Pramitarini, dan Novia Puspitasari, "Analisis Mengenai Calon Presiden Indonesia 2019 Di Twitter Menggunakan Metode Backpropagation," *J. Inform. Polinema*, vol. 6, no. 2, hal. 27–31, 2020, doi: 10.33795/jip.v6i2.306.
- [2] R. Ardiansyah, "Analisis Sentimen Calon Presiden Dan Wakil Presiden Periode 2019-2024 Pasca Debat Pilpres Di Twitter," *Sci. Comput. Sci. Informatics J.*, vol. 2, no. 1, hal. 21, 2019, doi: 10.22487/j26204118.2019.v2.i1.13068.
- [3] R. Tineges, A. Triayudi, dan I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, hal. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [4] N. Hendrastuty, A. Rahman Isnain, A. Yanti Rahmadhani, dan S. Styawati, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector

- Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, hal. 150–155, 2021, [Daring]. Tersedia pada: <http://situs.com>.
- [5] A. W. Wilantapoera, “Analisis Sentimen Kategori Aspek Pada Ulasan Produk Menggunakan Metode KNN Dengan Seleksi Fitur Mutual Information Program Studi Sarjana Informatika Fakultas Informatika Universitas Telkom Bandung Lembar Persetujuan Analisis Sentimen Kategori Aspek Pada U,” vol. 10, no. 2, hal. 1673–1681, 2022.
- [6] P. Widyantara dkk., “Analisis Sentimen pada Teks Berbahasa Bali Menggunakan Metode Multinomial Naive Bayes dengan TF-IDF dan BoW,” *Jnatia*, vol. 2, no. 1, hal. 37–46, 2023.
- [7] Merinda Lestandy, Abdurrahim Abdurrahim, dan Lailis Syafa’ah, “Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, hal. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.
- [8] T. F. T. Hidayat, G. Garno, dan A. A. Ridha, “Analisis Sentimen Opini Pemindahan Ibu Kota Pada Twitter Dengan Metode Support Vector Machine,” *J. Ilmu Komput.*, vol. 14, no. 1, hal. 49, 2021, doi: 10.24843/jik.2021.v14.i01.p06.
- [9] S. Al Qodrin dkk., “Klasifikasi Pertanyaan Berbahasa Indonesia Menggunakan Algoritma,” vol. 14, no. 2, hal. 44–52, 2022.
- [10] P. Arsi dan R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, hal. 147, 2021, doi: 10.25126/jtiik.0813944.
- [11] P. W. A. Wibawa dan C. Pramatha, “Systematic Literature Review: Machine Learning Methods in Emotion Classification in Textual Data,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 3, hal. 425–433, 2023, doi: 10.32736/sisfokom.v12i3.1787.
- [12] F. Rosyad, D. Pramono, dan K. C. Brata, “Analisis dan Perbaikan Usability Pada Aplikasi Ker Menggunakan Metode Usability Testing dan System Usability Scale (SUS),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 7, hal. 2261–2268, 2020.
- [13] M. L. L. Usman dan M. A. Gustalika, “Pengujian Validitas dan Reliabilitas System Usability Scale (SUS) Untuk Perangkat Smartphone,” *J. Ecotipe (Electronic, Control. Telecommun. Information, Power Eng.)*, vol. 9, no. 1, hal. 19–24, 2022, doi: 10.33019/jurnalecotipe.v9i1.2805.
- [14] A. S. ABADI dan P. F. DEWI, “Multimedia Mobile Application of National Heroes History Learning for Children’s Character Education,” *Telematika*, vol. 18, no. 3, hal. 308, 2021, doi: 10.31315/telematika.v18i3.5542.
- [15] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, dan B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM,” vol. 7, no. 1, hal. 11–18, 2022.
- [16] C. Pramatha, J. G. Davis, dan K. K. Y. Kuan, “A Semantically-Enriched Digital Portal for the Digital Preservation of Cultural Heritage with Community Participation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11196 LNCS, no. October, hal. 560–571, 2018, doi: 10.1007/978-3-030-01762-0_49.