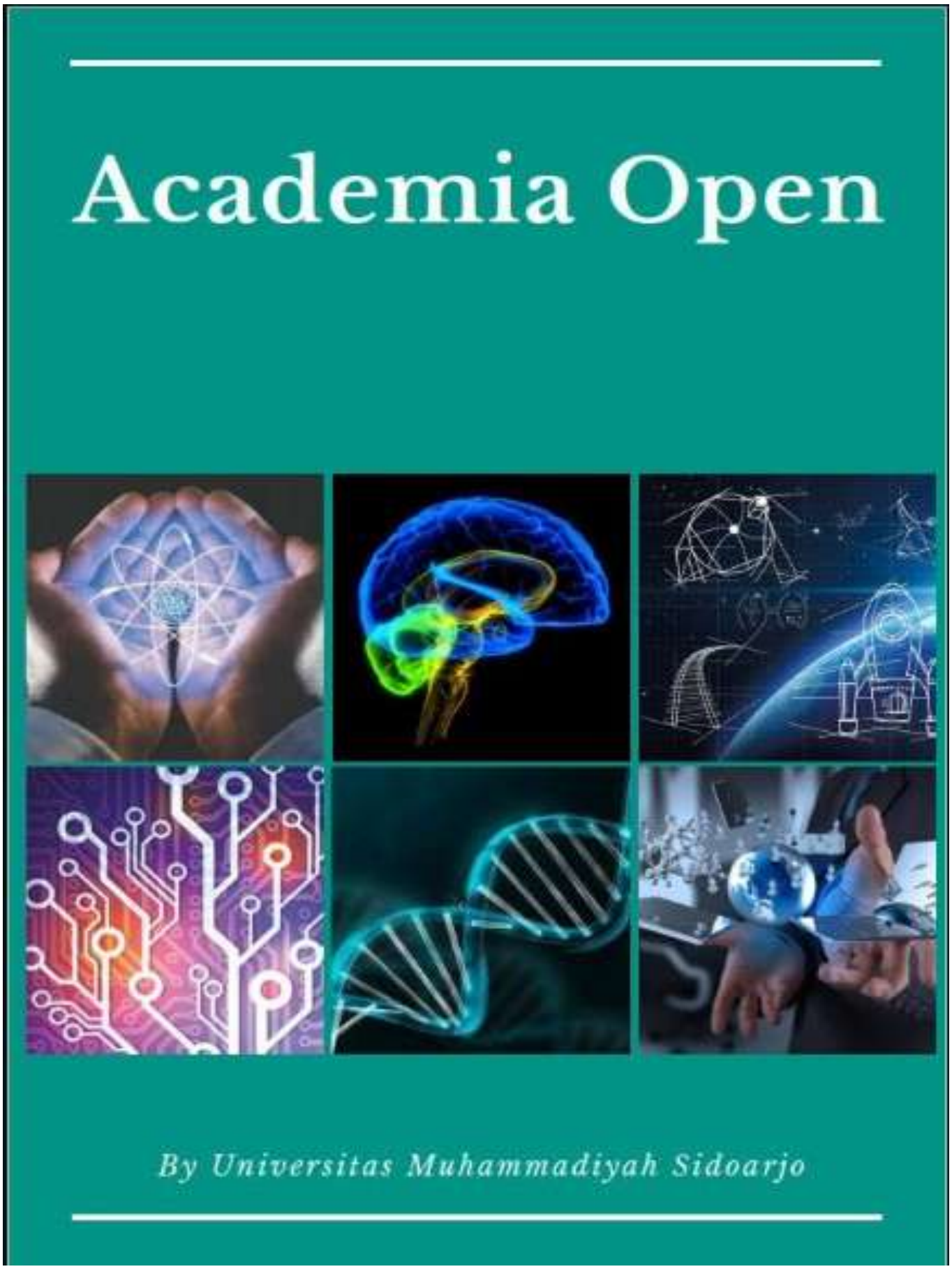
Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701



ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Table Of Contents

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of any other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright Statement

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

EDITORIAL TEAM

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Article information

Check this article update (crossmark)

Check this article impact (*)

Save this article to Mendeley

(*) Time for indexing process is various, depends on indexing database platform

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Random Forest Exceeds Support Vector Machine In Election Abstention Sentiment Analysis: Random Forest Melampaui Support Vector Machine Dalam Analisis Sentimen Golongan Putih

Kezia Cantika Saroinsong, keziasaroinsong07@gmail.com (*)
Program Studi Teknik Informatika, Universitas Negeri Manado, Indonesia

Irene R. H. T. Tangkawarow, irene.tangkawarow@unima.ac.id
Program Studi Teknik Informatika, Universitas Negeri Manado, Indonesia

Efraim R. S. Moningkey, fmoningkey@unima.ac.id
Program Studi Teknik Informatika, Universitas Negeri Manado, Indonesia

(*) Corresponding author

Abstract

General Background Public opinions regarding political participation are increasingly expressed on social media platforms. **Specific Background** During the 2024 Indonesian voting period, the white group phenomenon generated significant unstructured textual data on the X platform. **Knowledge Gap** Existing literature presents competitive capabilities between classification models, necessitating a targeted evaluation of their performance specifically on political non-voting discourse. **Aims** This study compares the classification capabilities of two predictive algorithms regarding public views on the 2024 white group movement. **Results** Utilizing 1,008 textual data points processed via term frequency-inverse document frequency, the decision tree-based model achieved higher metrics with 75% accuracy, 75% precision, 75% recall, and a 73% F1-score. Conversely, the hyperplane-based algorithm recorded 72% accuracy, 69% precision, 72% recall, and a 69% F1-score. Furthermore, data distribution revealed that 48.31% supported non-voting, 35.91% opposed it, and 15.78% remained neutral. **Novelty** The research uniquely integrates a Streamlit application that visually maps the entire machine learning pipeline alongside political distribution patterns. **Implications** These analytical findings provide institutions with critical insights to formulate targeted public communication strategies and encourage future citizen participation.

Highlights

- Predictive modeling demonstrated a 75% accuracy rate for identifying unstructured social media text patterns.
- Non-voting advocacy dominated the digital public discourse with 487 documented instances during the observation period.
- An interactive web application was successfully deployed to visualize the entire data processing workflow.

Keywords

Machine Learning; Text Classification; Public Opinion; Feature Extraction; Data Visualization

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Published date: 2026-09-15

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

PENDAHULUAN

Pemilihan umum merupakan salah wujud nyata demokrasi yang dilakukan rakyat sebagai perwujudan kehidupan tata negara yang demokratis. Pemilihan umum memberikan hak konstitusional bagi warga negara untuk menentukan arah kepemimpinan dan kebijakan publik [1]. Pemilu dianggap penting karena merupakan mekanisme yang mengatur pergantian atau perpindahan kekuasaan secara legal tanpa penggunaan kekerasan maupun cara-cara yang inkonstitusional [2]. Tetapi dalam pelaksanaannya banyak warga negara yang dengan sengaja tidak menggunakan hak pilihnya, fenomena ini dikenal dengan istilah, golongan putih. Secara historis, golongan putih tidaklah lahir dari orang-orang yang tidak paham akan pentingnya partisipasi masyarakat, dalam memberikan suaranya sebagai penentu masa depan bangsa, melainkan lahir dan dimotori oleh para intelektual yang motifnya untuk menghilangkan pemilu yang dipandang tidak sehat [3].

Teknologi sudah menjadi bagian dari hampir semua aktivitas manusia, salah satunya adalah perkembangan perangkat berbasis internet yang membuat berbagai informasi dapat dengan mudah diakses dari berbagai penjuru dunia. Di era digital saat ini, penyebaran informasi tidak selalu terkendali, sehingga berpotensi memunculkan ujaran kebencian yang dapat memengaruhi opini publik secara negatif [4]. Sentimen publik yang tersebar di berbagai platform dapat dengan mudah dianalisis karena penggunaan media sosial sering dijadikan wadah masyarakat dalam menyampaikan berbagai opini, salah satunya terkait politik. Seiring dengan itu, penerapan Machine Learning dalam analisis sentimen dapat membantu tugas manusia dengan mempermudah melakukan analisis terhadap banyaknya data yang tidak terstruktur dari media sosial yang sebelumnya membutuhkan waktu yang lama [5].

Klasifikasi dokumen teks ke dalam label netral, negatif, atau positif umumnya mengandalkan berbagai pendekatan analisis sentimen [6], Dua di antara algoritma yang paling populer digunakan adalah Random Forest dan Support Vector Machine. Kapabilitas Random Forest terletak pada ketangguhannya memproses data yang mengandung noise, kemampuannya mengelola memori secara efisien, serta ketahanannya dalam menjaga akurasi meski terdapat missing values [7]. Sementara itu, keunggulan Support Vector Machine dalam kategorisasi teks [8], bersumber pada mekanismenya yang membangun hyperplane optimal sebagai pemisah antar-kelas data [9], Hal ini menjadikan SVM sangat reliabel untuk data berdimensi tinggi dan mampu menghasilkan keputusan klasifikasi yang presisi walaupun ukuran data latihnya cenderung minim [10].

Penelitian yang dilakukan oleh Septiana dkk, 2024 [11] menunjukkan bahwa Support Vector Machine memiliki performa yang sedikit lebih unggul dibandingkan Random Forest, dengan selisih sekitar 2%. Sebaliknya, penelitian yang dilakukan oleh Sari dkk, 2024 [12] menunjukkan bahwa Random Forest memiliki performa yang sedikit lebih unggul dibandingkan Support Vector Machine dengan selisih sekitar 1%. Selanjutnya algoritma Random Forest dan Support Vector Machine memiliki kinerja yang setara dalam penelitian yang dilakukan oleh Dwinnie dkk, 2024 menggunakan data dari media sosial yang sifatnya tidak terstruktur [13]. Literatur-literatur yang ada sejalan dalam menyimpulkan bahwa kedua algoritma tersebut sering digunakan dalam analisis sentimen pada data teks tidak terstruktur yang diambil dari media sosial, terutama untuk isu-isu politik. Kedua algoritma diperlukan untuk membandingkan kinerjanya dalam studi kasus golongan putih pada pemilu 2024 dengan menggunakan data dari media sosial, karena beberapa penelitian terdahulu memperlihatkan kinerja dari algoritma Random Forest dan Support Vector Machine yang kompetitif.

Melalui pemanfaatan matrik penilaian yang mencakup Accuracy, Precision, Recall, dan F1-Score, algoritma Random Forest dan Support Vector Machine akan dianalisis kinerjanya dalam mengklasifikasikan arah opini masyarakat terkait fenomena golongan putih pada Pemilihan Umum 2024. Selain itu gambaran awal mengenai pola sentimen publik terhadap fenomena golongan putih pada pemilu 2024 menjadi salah satu hasil juga dari proses analisis pandangan Masyarakat yang dilakukan dalam penelitian ini. Serta akan diimplementasikan juga dashboard untuk menyajikan hasil analisis sentimen dan perbandingan kinerja model secara keseluruhan lewat visual yang ditampilkan, sehingga informasi yang dihasilkan dapat dipahami dengan lebih mudah dan informatif.

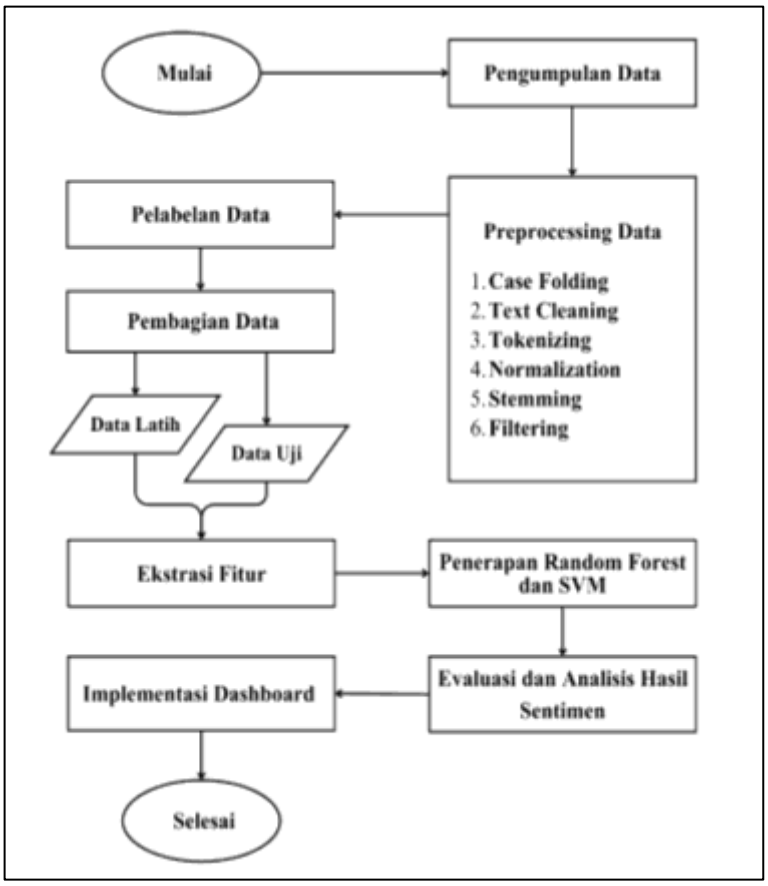
METODE

Klasifikasi pandangan publik mengenai fenomena golongan putih pada Pemilu 2024 diuji menggunakan dua model pembelajaran mesin, yaitu Random Forest dan Support Vector Machine, dengan memanfaatkan basis data sekunder. Fokus pengerjaan yang berpusat pada konversi dokumen teks ke dalam format numerik lewat ekstraksi fitur menjadi alasan diimplementasikannya metode kuantitatif komparatif. Lewat skema ini, evaluasi performa algoritma dapat dinilai secara objektif menggunakan analisis statistik yang valid. Di sisi lain, penggunaan pendekatan komparatif diaplikasikan untuk membandingkan hasil performa dari kedua algoritma melalui parameter ukur seperti Accuracy, Precision, Recall, dan F1-Score. Tahapan dari penelitian ini dirancang secara terstruktur guna menyajikan gambaran komprehensif mengenai langkah-langkah yang dilakukan selama proses penelitian berlangsung [14]. Tahapannya dimulai dari pengumpulan data dari media sosial, preprocessing data, pelabelan, pembagian, ekstrasi fitur, penerapan kedua algoritma hingga implementasi dashboard. Alur penelitiannya dapat dilihat pada Gambar 1 berikut ini.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Gambar 1 *Alur Penelitian*



A. Pengumpulan Data

Proses penelitian diawali dengan pengumpulan data relevan yang akan diolah menjadi sebuah informasi [15]. Data diambil dari media sosial X (sebelumnya Twitter) yang relevan dengan fenomena golongan putih. Melalui media Google Collaboratory dan pemanfaatan tool Tweet-Harvest, metode Web Crawling diaplikasikan untuk menghimpun data teks berbahasa Indonesia secara otomatis dari jejaring sosial X, yang merepresentasikan respons serta opini publik mengenai dinamika golongan putih pada Pemilu 2024. Digunakan rentang waktu yang sesuai dalam pengambilan data dalam penelitian ini yang memuat seluruh rangkaian kegiatan atau tahapan sebelum dilakukan pemilihan bahkan setelah dilakukan pemilihan umum yang terdiri dari tahapan awal kampanye, diskusi-diskusi politik dan arah pembangunan kedepan, hari pemungutan suara hingga setelah pemungutan suara hal ini guna menangkap opini, pandangan atau pembahasan masyarakat yang lebih akurat terhadap fenomena golongan putih pada pemilu 2024, oleh karena itu 01 Juni 2023 sampai 20 Februari 2024 dipilih sebagai rentang waktu pengambilan data dalam penelitian ini guna memastikan hasil evaluasi dan analisis dari penelitian ini bernilai representatif dan akurat dalam menangkap dinamika opini publik.

B. Preprocessing Data

Pembersihan data dari komponen tidak penting (seperti tautan, angka, dan simbol) lewat text cleaning, penyamaan format huruf menjadi kecil melalui case folding, serta pemecahan kalimat menjadi segmen kata individual lewat cara tokenizing merupakan bagian dari tahapan preprocessing. Prosedur ini menjadi kunci utama dalam membuat data mentah yang tidak terstruktur dan memiliki banyak noise menjadi data yang layak dan siap masuk ke tahap analisis [16]. Selanjutnya, dilakukan normalization untuk menyelaraskan kata slang atau singkatan ke bentuk baku, stemming untuk mereduksi kata menjadi bentuk dasarnya dengan melepas imbuhan, serta filtering guna menyaring kata-kata yang dianggap relevan dan signifikan bagi penelitian.

C. Pelabelan Data

Alur penelitian dilanjutkan ke fase penentuan label data setelah melalui tahapan preprocessing [17]. Tahapan pelabelan data dilakukan secara otomatis menggunakan pendekatan berbasis leksikon (lexicon-based approach) yang bekerja dengan cara mencocokkan kata-kata yang ada di dalam tweet dengan sebuah kamus sentimen (lexicon dictionary). Data akan dilabel menjadi 3 kelas yaitu kelas positif mendukung golongan putih (Pro-Golput), kelas negatif menolak golongan putih atau mengajak untuk memilih (Anti-Golput) serta kelas netral bersifat informatif atau tidak menunjukkan kecenderungan dukungan maupun penolakan terhadap fenomena golongan putih.

D. Pembagian Data

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Pembagian data dalam penelitian ini menggunakan rasio 80:20, yaitu 80% data digunakan sebagai data latih dan 20% sebagai data uji. Pemilihan rasio ini didasarkan pada praktik umum dalam penelitian Machine Learning yang bertujuan untuk menjaga keseimbangan antara proses pelatihan model dan evaluasi kinerja model. Beberapa penelitian terdahulu juga menerapkan rasio yang sama seperti yang dilakukan oleh Septiana dkk, 2024 [11] dalam analisis sentimen Quick Count Pemilu 2024 menggunakan pembagian data sebesar 80% untuk data latih dan 20% untuk data uji dalam proses pemodelan. Selain itu, penelitian oleh Rosanti dkk, 2023 [18] juga menerapkan pembagian data dengan rasio 80:20 dalam proses klasifikasi sentimen twitter terkait maskapai penerbangan. Penelitian lain oleh Tambuwun dkk, 2026 [19] juga menunjukkan penggunaan rasio 80:20 sebagai bagian dari proses pemodelan dan evaluasi kinerja algoritma.

E. Ekastrasi Fitur

Untuk menjembatani data teks yang telah bagi agar strukturnya dapat dikenali oleh algoritma machine learning teknik pembobotan kata seperti Term Frequency-Inverse Document Frequency (TF-IDF) diandalkan dalam penelitian ini untuk mengubah kata menjadi bentuk numerik sehingga dapat dimengerti oleh komputer (17). Konsep dasar TF-IDF mengarah pada penilaian bobot kepentingan sebuah kata pada suatu dokumen terhadap seluruh kumpulan data teks. Mekanisme pembobotannya menggabungkan dua komponen utama, yakni Term Frequency (TF) untuk melacak seberapa sering sebuah kata muncul yang dapat dirumuskan pada Persamaan 1, serta Inverse Document Frequency (IDF) untuk menilai bobot keunikan kata dalam skala dokumen yang lebih luas, dirumuskan pada Persamaan 2. Selanjutnya nilai TF-IDF diperoleh dengan mengalikan nilai keduanya, yang dirumuskan pada Persamaan 3 berikut ini.

$$TF(t,d) = \frac{Jumlah\ kemunculan\ term\ t\ dalam\ dokumen\ d}{Total\ jumlah\ term\ dalam\ dokumen\ d} \quad \text{Persamaan 1}$$

$$IDF\ (t,D) = log\ (\frac{Total\ jumlah\ dokumen\ dalam\ D}{Jumlah\ dokumen\ dalam\ D\ yang\ menggandung\ term\ t} + 1) \quad \text{Persamaan 2}$$

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

Dengan menggunakan metode TF-IDF, setiap kata dalam dokumen akan memiliki bobot tertentu yang mencerminkan tingkat kepentingannya, sehingga data teks dapat direpresentasikan dalam bentuk numerik yang siap digunakan dalam proses klasifikasi menggunakan algoritma Machine Learning [20].

F. Penerapan Random Forest

Metode Random Forest bekerja dengan cara membangun banyak pohon keputusan (decision trees) sekaligus secara bersamaan. Dalam proses pembentukannya, setiap pohon memilih karakteristik data secara acak untuk setiap node (titik percabangan). Hasil dari banyaknya pohon keputusan yang dibagun ini kemudian digabungkan untuk menentukan klasifikasi terhadap data yang digunakan. Proses penentuan kelas dalam algoritma Random Forest ini dapat dilihat melalui Persamaan 4 berikut ini.

$$f(x) = Average\left(f_1(x), f_2(x), \dots, f_n(x)\right) \quad \text{Persamaan 4}$$

Keterangan

$f(x)$	: Hasil prediksi
$f_1(x), f_2(x), \dots, f_n(x)$	: Hasil prediksi dari setiap pohon keputusan
x	: Data masukan

G. Penerapan Support Vector Machine

Algoritma Support Vector Machine banyak dipakai dalam mengklasifikasikan data teks dari berbagai media sosial ataupun dari berbagai sumber online, dimana algoritma ini memanfaatkan hyperplane atau pagar pembatas untuk mengklasifikasikan tiap data berdasarkan kelas masing-masing, algoritma ini tak hanya sebatas memanfaatkan hyperplane untuk mengklasifikasikan data yang linear, dimana ada kondisi masing-masing kelas terlalu bercampur, sehingga lumayan sulit untuk membangun hyperplane yang biasanya menggunakan kernel linear oleh karena itu Support Vector Machine dapat melempar data yang lainnya ke dimensi baru lalu membangun pagar atau hyperplane untuk membedakan data kedalam tiap kelas masing-masing, sehingga hasil klasifikasinya lebih akurat. Penyelesaian klasifikasi Support Vector Machine, dapat menggunakan persamaan berikut ini.

$(w \cdot x_i)+b=0$

Data x_i yang termasuk pada kelas -1 direpresentasikan dengan persamaan dibawah ini.

$(w \cdot x_i + b) \leq yi=-1$

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Untuk data x_i yang termasuk pada kelas +1, direpresentasikan dengan persamaan dibawah ini.

(w·x_i + b)≥ yi=1

Keterangan

w·x_i : Data ke – i.

x_i : Nilai bobot untuk kelas data ke – i.

b: Nilai bias.

yi: Kelas data ke – i.

H. Evaluasi dan Analisis Hasil Sentimen

Tahap akhir meliputi evaluasi dan analisis hasil sentimen, dengan menggunakan Classification Report yang mempresentasikan hasil seperti Accuracy, Precision, Recall, dan F1-Score, untuk menentukan algoritma mana yang memberikan performa terbaik dalam tugas klasifikasi sentimen. Perhitungan Classification Report adalah sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

.

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$Macro\ avg = \frac{M_1 + M_2 + \cdots + M_n}{n}$$

$$Weighted\ avg = \frac{(M_1 \times S_1) + (M_2 \times S_2) + \cdots + (M_n \times S_n)}{Total\ Support}$$

Keterangan

TP: Data yang diprediksi positif oleh model dan memang berlabel positif (True Positive).

FP: Data yang diprediksi positif oleh model, tetapi sebenarnya berlabel negative (False Positive).

TN: Data yang diprediksi negatif oleh model dan memang berlabel negative (True Negative).

FN: Data yang diprediksi negatif oleh model, tetapi sebenarnya berlabel positif (False Negative).

n : Jumlah total kelas (dalam penelitian ini n = 3, yaitu kelas positif, negatif dan netral).

M_n : Nilai metrik evaluasi spesifik pada kelas ke-n.

S_n : Jumlah sampel data asli (Support) pada kelas ke-n.

Total Support: Total keseluruhan sampel data uji dari seluruh kelas.

I. Implementasi Dashboard

Sebagai bentuk implementasi dari hasil penelitian, dikembangkan sebuah dashboard analisis sentimen berbasis Streamlit yang digunakan untuk memvisualisasikan seluruh tahapan penelitian. Dashboard tersebut diimplementasikan dalam bentuk aplikasi berbasis web, yang pada dasarnya merupakan kumpulan laman digital yang saling terintegrasi dan dapat diakses melalui jaringan internet [21].

HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Peneliti berhasil memperoleh sebanyak 1008 data yang memuat beberapa atribut seperti full text atau cuitan, username, dan informasi lainnya. Jumlah tersebut dinilai memadai untuk penelitian analisis sentimen berbasis Machine Learning, khususnya dalam penggunaan algoritma Random Forest dan Support Vector Machine. Hal ini didukung oleh beberapa penelitian terdahulu yang menunjukkan bahwa performa model klasifikasi tetap dapat optimal meskipun menggunakan

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

dataset dengan jumlah ratusan hingga ribuan data. Penelitian yang dilakukan oleh Sidauruk dkk, 2023 [22] menggunakan 655 data ulasan dari Google Playstore, memperoleh akurasi sebesar 93% pada Random Forest dan 97% pada Support Vector Machine. Selanjutnya penelitian yang dilakukan oleh Mursyid dkk, 2024 [23] menggunakan 806 data tweet opini masyarakat pasca pemilu presiden 2024 memperoleh akurasi sebesar 73,14% pada Random Forest dan 74,79% pada Support Vector Machine. Penelitian lain yang dilakukan oleh Septiana dkk, 2024 [11] menggunakan dataset yang lebih besar, yaitu 2000 data tweet terkait quick count pemilu 2024, menunjukkan bahwa kedua algoritma tetap mampu memberikan performa yang baik dengan akurasi sebesar 78% pada Random Forest dan 80% pada Support Vector Machine. Berdasarkan beberapa penelitian tersebut, penggunaan dataset dalam jumlah ratusan hingga ribuan data merupakan praktik yang umum dalam penelitian analisis sentimen dan mampu menghasilkan performa model yang baik. Oleh karena itu, jumlah sebanyak 1008 data dalam penelitian ini berada pada rentang yang representatif dan dinilai cukup untuk menggambarkan pola sentimen publik terhadap fenomena golongan putih pada pemilu 2024.

B. Hasil Preprocessing Data

Tahapan preprocessing data telah dilakukan terhadap 1.008 data tweet mentah, yang prosesnya terdiri dari case folding, text cleaning, tokenizing, normalization, stemming dan filtering untuk menghasilkan data bersih dan siap diolah. Proses yang dilakukan dapat dilihat pada Tabel 1 berikut ini.

Tabel 1 *Hasil Tahapan Preprocessing*

Tahapan	**Data (Tweet)**
Data Mentah	@OposisiCerdas jika sudah selesai dan resmi amin kalah dan Koalisi amin bergabung dengan si pemenang ini pemilu berikutnya jika di berikan umur lebih baik Golput saja
Case Folding	@oposisicerdas jika sudah selesai dan resmi amin kalah dan koalisi amin bergabung dengan si pemenang ini pemilu berikutnya jika di berikan umur lebih baik golput saja
Text Cleaning	jika sudah selesai dan resmi amin kalah dan koalisi amin bergabung dengan si pemenang ini pemilu berikutnya jika di berikan umur lebih baik golput saja
Tokenizing	['jika', 'sudah', 'selesai', 'dan', 'resmi', 'amin', 'kalah', 'dan', 'koalisi', 'amin', 'bergabung', 'dengan', 'si', 'pemenang', 'ini', 'pemilu', 'berikutnya', 'jika', 'di', 'berikan', 'umur', 'lebih', 'baik', 'golput', 'saja']
Normalization	['jika', 'sudah', 'selesai', 'dan', 'resmi', 'amin', 'kalah', 'dan', 'koalisi', 'amin', 'bergabung', 'dengan', 'sih', 'pemenang', 'ini', 'pemilu', 'berikutnya', 'jika', 'di', 'berikan', 'umur', 'lebih', 'baik', 'golput', 'saja']
Stemming	['jika', 'sudah', 'selesai', 'dan', 'resmi', 'amin', 'kalah', 'dan', 'koalisi', 'amin', 'gabung', 'dengan', 'sih', 'menang', 'ini', 'milu', 'ikut', 'jika', 'di', 'ikan', 'umur', 'lebih', 'baik', 'golput', 'saja']
Filtering	['selesai', 'resmi', 'amin', 'kalah', 'koalisi', 'amin', 'gabung', 'menang', 'pemilu', 'ikut', 'ikan', 'umur', 'baik', 'golput']

Berdasarkan Tabel 1, dapat dilihat bahwa panjang teks dan keberagaman karakter telah tereduksi secara signifikan tanpa menghilangkan konteks utama dari kalimat. Kumpulan token bersih inilah yang merepresentasikan fitur dari setiap tweet.

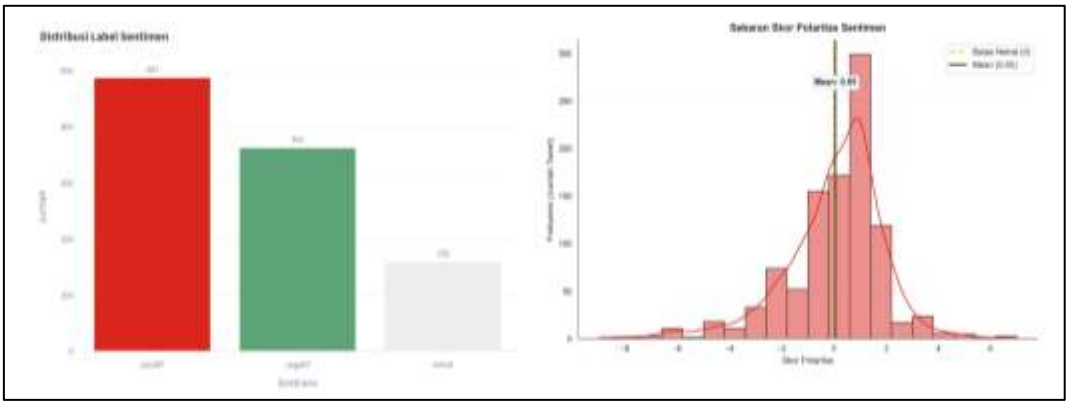
C. Hasil Pelabelan Data

Dalam pelabelan data kamus sentimen yang dipakai terdiri dari dua sumber, yaitu kamus sentimen eksternal yang berisi daftar kata positif dan negatif, lalu kamus sentimen kustom yang dibuat secara spesifik oleh peneliti untuk menangkap konteks fenomena golongan putih. Data dilabel menjadi 3 kelas yaitu kelas positif mendukung golongan putih, negatif menolak golongan putih serta netral bersifat informatif atau tidak menunjukkan kecenderungan dukungan dukungan maupun penolakan terhadap fenomena golongan putih. Hasil pelabelan data dapat dilihat pada Gambar 2 berikut ini.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Gambar 2 *Distribusi Label dan Sebaran Skor Polaritas Sentimen*



Berdasarkan Gambar 2, dapat diamati bahwa sentimen Positif (Pro-Golput) mendominasi percakapan di media sosial dengan jumlah 487 data (48,31%). Hal ini menunjukkan bahwa banyak masyarakat yang secara terbuka mengekspresikan kekecewaan, menyuarakan hak untuk tidak memilih, atau memaklumi tindakan golput pada pemilu 2024. Di sisi lain, 362 data (35,91%) bersentimen Negatif (Anti-Golput) yang berisi narasi perlawanan terhadap golput dan ajakan untuk memilih, sementara sisanya 159 data (15,78%) bersifat Netral. Temuan tersebut juga diperkuat oleh sebaran skor polaritas dimana sumbu horizontal merepresentasikan nilai skor polaritas, sedangkan sumbu vertikal menunjukkan frekuensi jumlah data. Garis vertikal merah putus-putus menandai batas netral (skor = 0), yang digunakan sebagai acuan dalam klasifikasi sentimen, di mana skor di atas 0 dikategorikan sebagai positif, di bawah 0 sebagai negatif, dan 0 sebagai netral. Penelitian ini juga mengamati kata-kata yang paling dominan muncul pada masing-masing kategori, ditunjukkan pada Gambar 3 berikut ini.

Gambar 3 *World Cloud Positif, Negatif dan Netral*

Gambar 3 menunjukkan perbedaan karakteristik kata pada masing-masing kelas sentimen. Kata-kata dalam kategori positif seperti "golput", "tidak", "dukung", dan "benar" memperlihatkan orang-orang yang mendukung tindakan golongan putih yang memilih untuk golput dan memaklumi tindakan ini, lalu kata-kata dalam kategori negatif seperti "pilih", "nyoblos", "tidak", dan "suara" memperlihatkan masyarakat banyak yang menyuarakan untuk menggunakan hak pilih dan jangan golput sedang dalam kategori netral kata-kata dominan seperti "pilih", "golput", "kalau", dan "orang" memperlihatkan pembahasan yang mengarah ke penyampaian informasi tanpa mendukung maupun menolak tindakan golongan putih. Langkah berikutnya setelah proses distribusi label selesai adalah membagi seluruh kumpulan data yang telah diberi label ke dalam pembagian data yaitu training data untuk melatih atau mengajari algoritma supaya pintar dan testing data untuk melihat atau mengevaluasi kemampuan algoritma yang telah diajari atau dilatih sebelumnya dalam mengklasifikasikan data.

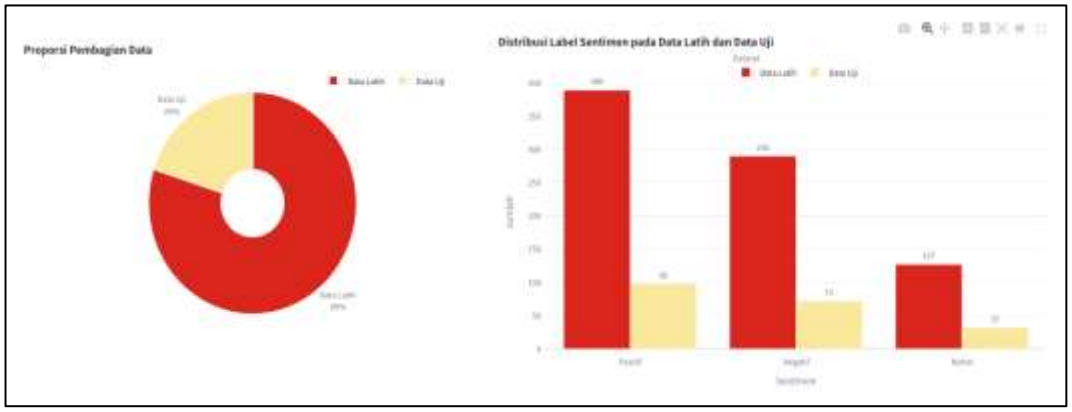
D. Hasil Pembagian Data

Pembagian data dalam penelitian ini menggunakan rasio 80:20, yaitu 80% data digunakan sebagai data latih dan 20% sebagai data uji. Hasil pembagian data divisualisasikan pada Gambar 4 berikut ini.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

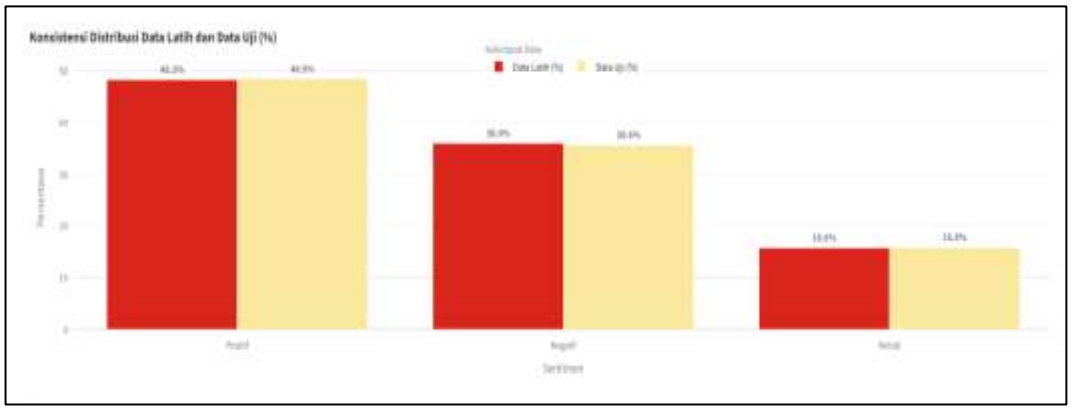
Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Gambar 4 *Hasil Pembagian Data*



Dari total 1008 data tweet, sebanyak 806 data (80%) digunakan sebagai data latih, sedangkan 202 data (20%) digunakan sebagai data uji. Penelitian ini juga menerapkan teknik stratifikasi yang berfungsi untuk mengunci dan menduplikasi rasio persentase setiap kelas dari populasi awal agar tetap sama dan konsisten, baik di dalam subset data latih maupun subset data uji. Pada data latih terdapat 389 data positif, 290 data negatif, dan 127 data netral, sedangkan pada data uji terdapat 98 data positif, 72 data negatif, dan 32 data netral. Distribusi ini menunjukkan bahwa proporsi masing-masing kelas relatif seimbang antara data latih dan data uji. Hal ini penting untuk memastikan bahwa model tidak bias terhadap salah satu kelas tertentu. Untuk melihat lebih jelas tingkat konsistensi distribusi label, hasilnya divisualisasikan pada Gambar 5 berikut ini.

Gambar 5 *Konsistensi Distribusi Label Data Latih dan Data Uji*



Berdasarkan visualisasi pada Gambar 5, dapat dikonfirmasi bahwa teknik stratifikasi telah berhasil menjaga keseimbangan distribusi data secara proporsional. Sebagai contoh, sentimen positif memiliki persentase sekitar 48,3% pada data latih dan 48,5% pada data uji. Konsistensi ini memastikan bahwa algoritma Random Forest dan Support Vector Machine akan menerima porsi pembelajaran kelas yang adil.

E. Hasil Ekstraksi Fitur TF-IDF

Ekstraksi fitur TF-IDF dilakukan menggunakan modul TfidfVectorizer dari pustaka Scikit-Learn. Untuk mengoptimalkan kinerja memori dan mencegah model dari kondisi overfitting. Salah satu prosedur terpenting dalam tahapan ini adalah pencegahan kebocoran informasi (data leakage). Proses pembelajaran kosa kata dan pembobotan (fit_transform) hanya dilakukan pada data latih (806 data). Sedangkan untuk data uji (202 data), hanya dilakukan proses transformasi (transform) berdasarkan kamus kata yang telah dipelajari dari data latih agar data uji tetap bersifat unseen data. Dimensi matriks TF-IDF yang berhasil dibentuk ditunjukan pada Gambar 6 berikut ini.

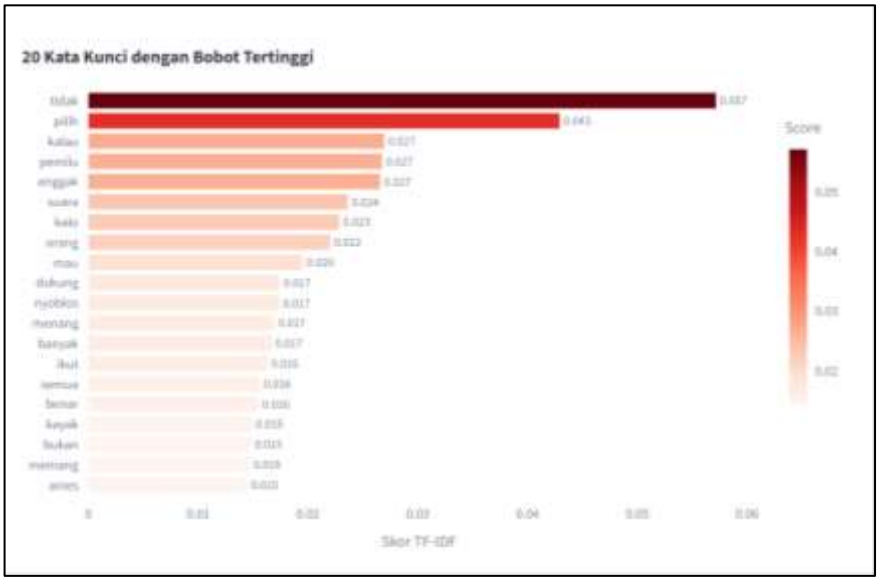
Gambar 6 *Hasil Dimensi Matriks TF-IDF*

```
TF-IDF berhasil dihitung
Dimensi TF-IDF data latih: (806, 986)
Dimensi TF-IDF data uji: (202, 986)
```

Hasil dimensi ini menunjukkan bahwa dari ribuan kata acak, ekstraksi fitur berhasil menyeleksi 986 fitur kata terbaik yang memenuhi kriteria. Setiap tweet kini telah direpresentasikan ke dalam bentuk matriks dengan 986 kolom fitur. Visualisasi 20 fitur kata (top 20 features) dengan bobot rata-rata tertinggi disajikan pada Gambar 7 berikut ini.

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Gambar 7 *Top 20 Fitur (Kata) Berdasarkan Bobot TF-IDF Tertinggi*



Gambar 7 menampilkan 20 kata dengan nilai rata-rata skor TF-IDF tertinggi yang dihasilkan dari proses ekstraksi fitur. Terdapat kata tidak, pilih, dan kalau yang menempati posisi teratas lalu diikuti kata pemilu, enggak dan suara yang banyak dipakai dalam pembahasan lalu ada kata nyoblos, dukung yang mencerminkan pembahasan yang sesuai dengan studi kasus penelitian yaitu fenomena golongan putih pada pemilu 2024. Kata-kata beserta bobot numerik inilah yang kemudian dijadikan sebagai masukan atau input utama bagi algoritma klasifikasi Random Forest dan Support Vector Machine untuk mengidentifikasi kelas sentimen.

F. Hasil Klasifikasi Menggunakan Random Forest

Digunakan parameter n_estimators=100 yang berarti membangun sebanyak 100 pohon keputusan untuk mengklasifikasikan tiap data yang ada, serta agar menjaga konsistensi dari hasil yang didapatkan maka random_state=42 digunakan supaya setiap kali dijalankan kodenya, hasil pengacakan data untuk masing-masing pohon akan selalu sama. 986 fitur dari perolehan ekstrasi TF-IDF dan 806 data digunakan untuk melatih model dan 202 untuk menguji model. Hasil evaluasi model Random Forest menunjukkan nilai akurasi sebesar 0,75 atau 75% yang ditunjukan pada Gambar 8 berikut ini.

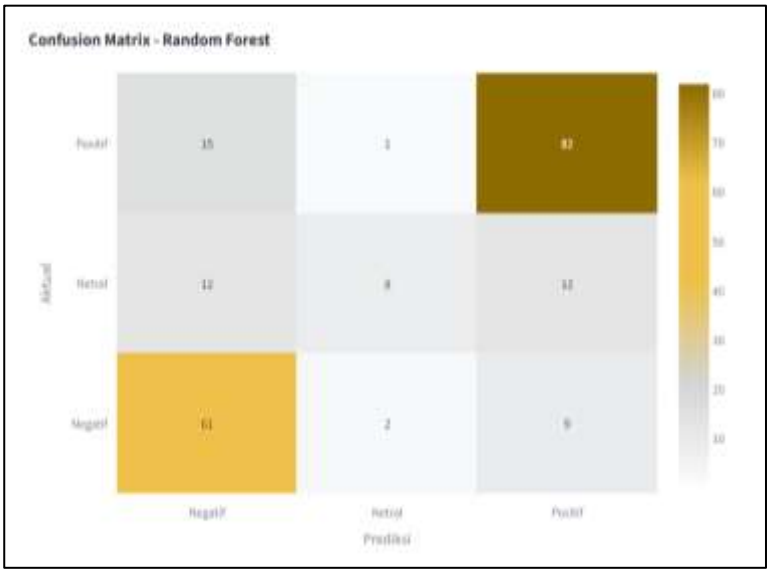
Gambar 8 *Hasil Classification Report Random Forest*

```
HASIL EVALUASI RF
Akurasi: 0.75

              precision    recall  f1-score   support

     negatif       0.69      0.85      0.76        72
      netral       0.73      0.25      0.37        32
     positif       0.80      0.84      0.82        98

    accuracy                           0.75       202
   macro avg       0.74      0.64      0.65       202
weighted avg       0.75      0.75      0.73       202
```

Evaluasi model berdasarkan Classification Report memperlihatkan bahwa tingkat keberhasilan klasifikasi berbeda pada setiap kategori sentimen. Kelas Positif (Pro-Golput) memperoleh hasil paling baik dengan nilai precision 0,80, recall 0,84, dan f1-score 0,82, yang menunjukkan kemampuan model dalam mengenali serta memprediksi sentimen positif secara konsisten. Pada kelas Negatif (Anti-Golput), model mencatat recall tertinggi sebesar 0,85 dengan precision 0,69 dan f1-score 0,76, mengindikasikan bahwa sebagian besar data negatif berhasil teridentifikasi meskipun masih terdapat sejumlah kesalahan klasifikasi. Sementara itu, kategori Netral menjadi kelas yang paling sulit diprediksi dengan recall hanya 0,25, meskipun precision mencapai 0,73 dan f1-score sebesar 0,37. Rendahnya performa pada kelas ini mengindikasikan bahwa karakteristik sentimen netral cenderung kurang memiliki pola yang tegas sehingga sering kali memiliki kemiripan dengan sentimen positif maupun negatif. Untuk memperoleh pemahaman yang lebih mendalam mengenai distribusi prediksi dan pola kesalahan klasifikasi yang dihasilkan model, analisis selanjutnya dilakukan melalui Confusion Matrix yang disajikan pada Gambar 9.

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Gambar 9 *Confusion Matrix Random Forest*



Berdasarkan pemetaan pada diagram Confusion Matrix, sistem sukses memetakan data secara presisi sebanyak 61 sampel untuk kelas negatif, 8 sampel pada kelas netral, dan 82 sampel di kelompok positif. Adapun rincian galat (error) prediksi yang terjadi dimana pada kelompok negatif, terdapat 2 data yang meleset ke dalam kategori netral dan 9 data teridentifikasi sebagai positif. Disini kelas netral yang dihasilkan dari algoritma random forest mengalami kesulitan dalam membedakan kategori yaitu masing-masing ada 12 data yang salah diklasifikasikan kedalam kelas positif dan netral, pada kelas positif 15 data salah masuk ke kelas negatif dan 1 data ke kelas netral. Untuk membuktikan nilai evaluasi yang tertera dalam Classification Report Random Forest pada Gambar 8, dilakukan perhitungan manual untuk nilai Accuracy, Precision, Recall, F1-Score, Macro Average dan Weighted Average berdasarkan data dari Confusion Matrix pada Gambar 9, perhitungannya adalah sebagai berikut.

$$\text{Kelas positif} = \frac{82}{82+15+1} = \frac{82}{98} = 0{,}8367\ldots \approx 0{,}84$$

4. F1-Score

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Kelas negatif} = 2 \times \frac{0{,}69 \times 0{,}85}{0{,}69 + 0{,}85} = \frac{0{,}5865}{1{,}54} = 0{,}76$$

$$\text{Kelas netral} = 2 \times \frac{0{,}73 \times 0{,}25}{0{,}73 + 0{,}25} = \frac{0{,}1825}{0{,}98} = 0{,}3724\ldots \approx 0{,}37$$

$$\text{Kelas positif} = 2 \times \frac{0{,}80 \times 0{,}84}{0{,}80 + 0{,}84} = \frac{0{,}672}{1{,}64} = 0{,}8195\ldots \approx 0{,}82$$

5. Macro Average

$$\text{Macro avg} = \frac{M_1 + M_2 + \cdots + M_n}{n}$$

$$\text{Precision} = \frac{0{,}69 + 0{,}73 + 0{,}80}{3} = \frac{2{,}22}{3} = 0{,}74$$

$$\text{Recall} = \frac{0{,}85 + 0{,}25 + 0{,}84}{3} = \frac{1{,}94}{3} = 0{,}64$$

$$\text{F1-score} = \frac{0{,}76 + 0{,}37 + 0{,}82}{3} = \frac{1{,}95}{3} = 0{,}65$$

6. Weighted Average

$$\text{Weighted avg} = \frac{(M_1 \times S_1) + (M_2 \times S_2) + \cdots + (M_n \times S_n)}{\text{Total Support}}$$

$$\text{Precision} = \frac{(0{,}69 \times 72) + (0{,}73 \times 32) + (0{,}80 \times 98)}{202} = \frac{49{,}68 + 23{,}36 + 78{,}4}{202} = \frac{151{,}44}{202} = 0{,}7497 \approx 0{,}75$$

$$\text{Recall} = \frac{(0{,}85 \times 72) + (0{,}25 \times 32) + (0{,}84 \times 98)}{202} = \frac{61{,}2 + 8{,}0 + 82{,}32}{202} = \frac{151{,}52}{202} = 0{,}7500 \approx 0{,}75$$

$$\text{F1-score} = \frac{(0{,}76 \times 72) + (0{,}37 \times 32) + (0{,}82 \times 98)}{202} = \frac{54{,}72 + 11{,}84 + 80{,}36}{202} = \frac{146{,}92}{202} = 0{,}7273 \approx 0{,}73$$

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

1. Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$
$$= \frac{61 + 8 + 82}{202} = \frac{151}{202} = 0{,}7475... \approx 0{,}75$$

2. Precision

$$\text{Precision} = \frac{TP}{TP+FP}$$
$$\text{Kelas negatif} = \frac{61}{61+12+15} = \frac{61}{88} = 0{,}6931... \approx 0{,}69$$
$$\text{Kelas netral} = \frac{8}{8+2+1} = \frac{8}{11} = 0{,}7272... \approx 0{,}73$$
$$\text{Kelas positif} = \frac{82}{82+12+9} = \frac{82}{103} = 0{,}7961... \approx 0{,}80$$

3. Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$
$$\text{Kelas negatif} = \frac{61}{61 + 2 + 9} = \frac{61}{72} = 0{,}8472... \approx 0{,}85$$
$$\text{Kelas netral} = \frac{8}{8 + 12 + 12} = \frac{8}{32} = 0{,}25$$

G. Hasil Klasifikasi Menggunakan Support Vector Machine

Support Vector Machine diimplementasikan menggunakan pustaka scikit-learn dengan kernel linear dan parameter random_state=42 untuk menjaga konsistensi hasil. Model dilatih menggunakan 806 data latih dengan 986 fitur hasil ekstraksi TF-IDF. Selanjutnya, model diuji menggunakan 202 data uji. Berdasarkan hasil pengujian, model SVM memperoleh nilai akurasi sebesar 0,72 atau 72% ditunjukan pada Gambar 10 berikut.

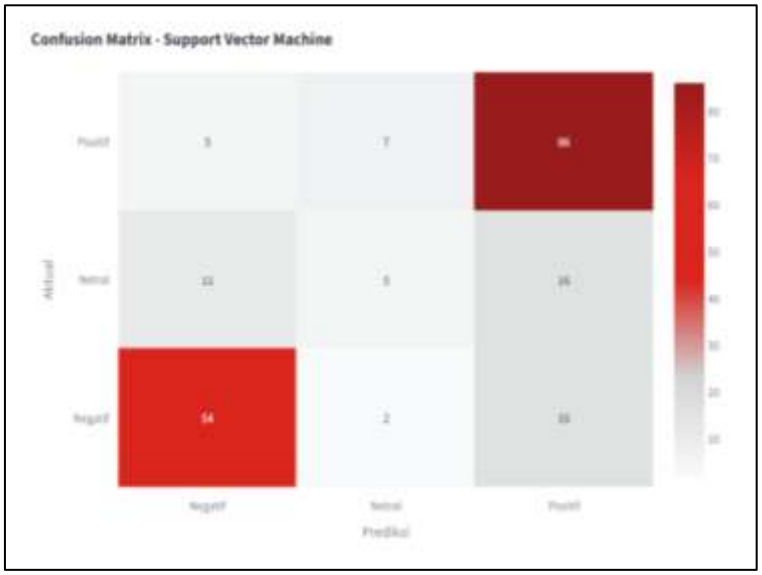
Gambar 10 *Hasil Classification Report Support Vector Machine*

```
HASIL EVALUASI SVM
Akurasi: 0.72

              precision    recall  f1-score   support

     negatif       0.77      0.75      0.76        72
      netral       0.36      0.16      0.22        32
     positif       0.73      0.88      0.80        98

    accuracy                           0.72       202
   macro avg       0.62      0.59      0.59       202
weighted avg       0.69      0.72      0.69       202
```

Berdasarkan Classification Report, performa model menunjukkan variasi pada masing-masing kelas sentimen. Pada kelas Positif (Pro-Golput), model memperoleh nilai recall tertinggi sebesar 0,88 dengan precision sebesar 0,73 dan f1-score sebesar 0,80. Hal ini menunjukkan bahwa model sangat baik dalam mengenali sebagian besar data positif, meskipun masih terdapat kesalahan prediksi terhadap kelas lain. Pada kelas Negatif (Anti-Golput), model menunjukkan performa yang cukup seimbang dengan nilai precision sebesar 0,77, recall sebesar 0,75, dan f1-score sebesar 0,76. Hal ini menunjukkan bahwa model cukup stabil dalam mengklasifikasikan sentimen negatif. Sebaliknya, performa yang rendah kembali terjadi pada kelas Netral, dengan nilai precision sebesar 0,36, recall sebesar 0,16, dan f1-score sebesar 0,22. Nilai recall yang rendah menunjukkan bahwa model mengalami kesulitan yang lebih besar dalam membentuk batas pemisah yang jelas untuk kelas netral. Hal tersebut diduga karena karakteristik tweet netral memiliki representasi kata yang cenderung beririsan dengan kelas positif maupun negatif setelah diekstraksi menggunakan TF-IDF, sehingga banyak data netral lebih mudah diprediksi sebagai salah satu dari dua kelas tersebut. Untuk melihat distribusi hasil prediksi secara lebih rinci, digunakan Confusion Matrix yang ditunjukkan pada Gambar 11 berikut ini.

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Gambar 11 *Confusion Matrix Support Vector Machine*



Berdasarkan Confusion Matrix, model berhasil mengklasifikasikan dengan benar sebanyak 54 data negatif, 5 data netral, dan 86 data positif. Pada kelas negatif, terdapat 2 data yang salah diprediksi sebagai netral dan 16 data sebagai positif. Pada kelas netral, sebagian besar data salah diklasifikasikan sebagai negatif 11 data dan positif 16 data. Sementara itu, pada kelas positif terdapat 5 data yang salah diprediksi sebagai negatif dan 7 data sebagai netral. Untuk membuktikan nilai evaluasi yang tertera dalam Classification Report Support Vector Machine pada Gambar 10, dilakukan perhitungan manual untuk nilai Accuracy, Precision, Recall, F1-Score, Macro Average dan Weighted Average berdasarkan data dari Confusion Matrix pada Gambar 11, perhitungannya adalah sebagai berikut.

1. Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$
$$= \frac{54 + 5 + 86}{202} = \frac{145}{202} = 0{,}7178\ldots \approx 0{,}72$$

2. Precision

$$\text{Precision} = \frac{TP}{TP+FP}$$
$$\text{Kelas negatif} = \frac{54}{54+11+5} = \frac{54}{70} = 0{,}7714\ldots \approx 0{,}77$$
$$\text{Kelas netral} = \frac{5}{5+2+7} = \frac{5}{14} = 0{,}3571\ldots \approx 0{,}36$$
$$\text{Kelas positif} = \frac{86}{86+16+16} = \frac{86}{118} = 0{,}7288\ldots \approx 0{,}73$$

3. Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

$$\text{Kelas negatif} = \frac{54}{54+2+16} = \frac{54}{72} = 0{,}75$$
$$\text{Kelas netral} = \frac{5}{5+11+16} = \frac{5}{32} = 0{,}1562... \approx 0{,}16$$
$$\text{Kelas positif} = \frac{86}{86+5+7} = \frac{86}{98} = 0{,}8775... \approx 0{,}88$$

4. F1-Score

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
$$\text{Kelas negatif} = 2 \times \frac{0{,}77 \times 0{,}75}{0{,}77+0{,}75} = \frac{0{,}5775}{1{,}52} = 0{,}7598... \approx 0{,}76$$
$$\text{Kelas netral} = 2 \times \frac{0{,}36 \times 0{,}16}{0{,}36+0{,}16} = \frac{0{,}0576}{0{,}52} = 0{,}2215... \approx 0{,}22$$
$$\text{Kelas positif} = 2 \times \frac{0{,}73 \times 0{,}88}{0{,}73+0{,}88} = \frac{0{,}6424}{1{,}61} = 0{,}7980... \approx 0{,}80$$

5. Macro Average

$$\text{Macro avg} = \frac{M_1 + M_2 + \cdots + M_n}{n}$$
$$\text{Precision} = \frac{0{,}77+0{,}36+0{,}73}{3} = \frac{1{,}86}{3} = 0{,}62$$
$$\text{Recall} = \frac{0{,}75+0{,}16+0{,}88}{3} = \frac{1{,}79}{3} = 0{,}5966... \approx 0{,}59$$
$$\text{F1-score} = \frac{0{,}76+0{,}22+0{,}80}{3} = \frac{1{,}78}{3} = 0{,}5933... \approx 0{,}59$$

6. Weighted Average

$$\text{Weighted avg} = \frac{(M_1 \times S_1) + (M_2 \times S_2) + \cdots + (M_n \times S_n)}{\text{Total Support}}$$
$$\text{Precision} = \frac{(0{,}77 \times 72) + (0{,}36 \times 32) + (0{,}73 \times 98)}{202} = \frac{55{,}44 + 11{,}52 + 71{,}54}{202} = \frac{138{,}5}{202} = 0{,}6856 \approx 0{,}69$$
$$\text{Recall} = \frac{(0{,}75 \times 72) + (0{,}16 \times 32) + (0{,}88 \times 98)}{202} = \frac{54{,}0 + 5{,}12 + 86{,}24}{202} = \frac{145{,}36}{202} = 0{,}7196 \approx 0{,}72$$
$$\text{F1-Score} = \frac{(0{,}76 \times 72) + (0{,}22 \times 32) + (0{,}80 \times 98)}{202} = \frac{54{,}72 + 7{,}04 + 78{,}4}{202} = \frac{140{,}16}{202} = 0{,}6938 \approx 0{,}69$$

H. Analisis Komparatif Kinerja Random Forest dan Support Vector Machine

Untuk mengetahui model klasifikasi yang paling optimal dalam penelitian ini, dilakukan analisis komparatif terhadap kinerja algoritma Random Forest dan Support Vector Machine. Perbandingan dilakukan berdasarkan empat metrik evaluasi utama, yaitu Accuracy, Precision, Recall, dan F1-Score yang diperoleh. Hasil perbandingan kinerja kedua algoritma tersebut ditunjukan pada Tabel 2 dan Gambar 12 berikut ini.

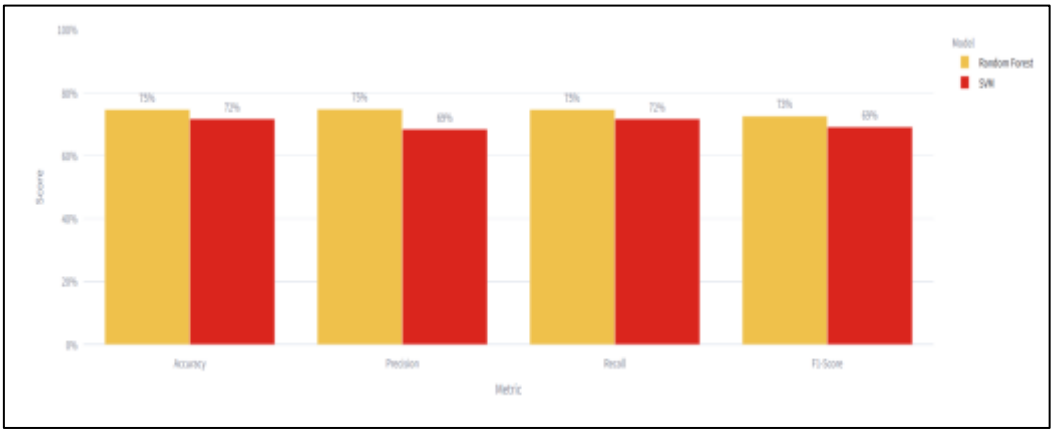
Tabel 2 *Evaluasi Komparatif Random Forest dan Support Vector Machine*

	Random Forest	Support Vector Machine	Selisih
Accuracy (%)	75	72	3
Precision (%)	75	69	6
Recall (%)	75	72	3
F1-Score (%)	73	69	4

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

Gambar 12 *Evaluasi Komparatif Random Forest dan Support Vector Machine*



Berdasarkan hasil evaluasi kinerja yang dilakukan menggunakan empat metrik utama, yaitu Accuracy, Precision, Recall, dan F1-Score. Algoritma Random Forest menunjukkan performa yang sedikit lebih unggul dibandingkan Support Vector Machine pada seluruh metrik evaluasi. Dengan selisih sebesar 3% yang tercatat pada metrik Accuracy, di mana Random Forest menyentuh angka 75% sementara Support Vector Machine menyentuh angka 72%. Untuk metrik Precision, efektivitas Random Forest 75% melampaui Support Vector Machine 69% dengan perbedaan sebesar 6%, yang mengisyaratkan ketajaman Random Forest dalam menekan bias salah klasifikasi pada label yang tidak relevan. Kemampuan dalam mengenali data aktual dari setiap kelas lewat metrik Recall diungguli oleh Random Forest juga dengan nilai 75%, dimana nilai 72% diperoleh dari Support Vector Machine sehingga memiliki perbedaan selisih 3%. Keseimbangan Precision dan Recall ada pada metrik F1-Score, dimana 73% diperoleh dari Random Forest dan 69% dari Support Vector Machine sehingga pada metrik ini Random Forest juga mengungguli Support Vector Machine dengan menghasilkan perbedaan performa akhir sebesar 4%. Hal ini menunjukkan bahwa secara keseluruhan, kedua algoritma memiliki performa yang kompetitif dalam melakukan klasifikasi terhadap pandangan masyarakat terkait fenomena golongan putih pada pemilu 2024.

I. Hasil Implementasi Dashboard Analisis Sentimen Berbasis Streamlit

Sebagai bentuk implementasi dari hasil penelitian, dikembangkan sebuah dashboard analisis sentimen berbasis Streamlit yang menyajikan seluruh tahapan penelitian meliputi proses pengumpulan data, preprocessing data, pelabelan data, pembagian data, ekstraksi fitur, evaluasi model, hingga kesimpulan yang dirancang untuk memudahkan pengguna dalam memahami alur penelitian. Visualisasinya ditunjukan pada Gambar 13 berikut ini.

Gambar 13 *Bagian Utama Dashboard*

Penyajian informasi dalam bentuk visual yang interaktif dan multimodal dapat membantu meningkatkan pemahaman terhadap data yang disajikan [24]. Visualisasi tersebut membantu dalam mengidentifikasi kecenderungan sentimen publik, kata-kata dominan, serta performa masing-masing algoritma. Implementasi dashboard ini memberikan kemudahan bagi pengguna dengan memberikan gambaran komprehensif, yang memuat visualisasi seperti grafik distribusi sentimen, histogram skor polaritas, word cloud, serta perbandingan kinerja model, sehingga pengguna tidak hanya melihat hasil akhir, tetapi juga memahami bagaimana hasil tersebut diperoleh dengan jelas.

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Random Forest dan Support Vector Machine sama-sama memiliki kemampuan yang baik dalam melakukan klasifikasi sentimen pada data teks. Meskipun Random Forest menunjukkan performa yang sedikit lebih unggul, kedua algoritma tetap memberikan hasil yang kompetitif sehingga pemilihannya dapat disesuaikan dengan kebutuhan dan karakteristik data. Selain itu, penelitian ini tidak hanya menghasilkan perbandingan kinerja algoritma dan visualisasi hasil melalui dashboard yang mendukung interpretasi data secara lebih efektif, tetapi juga memberikan gambaran awal kepada pemangku kepentingan atau lembaga terkait mengenai kecenderungan sentimen terlebih sentimen positif atau pro golput yang berkembang di media sosial yang dapat dimanfaatkan sebagai bahan evaluasi untuk merancang strategi komunikasi publik yang lebih efektif, memperkuat edukasi mengenai pentingnya partisipasi dalam pemilu, serta mengidentifikasi faktor-faktor yang mendorong munculnya kecenderungan golput dalam upaya peningkatan partisipasi pemilih pada pemilu berikutnya. Meskipun begitu penggunaan data dalam penelitian ini perlu dipahami berdasarkan ruang lingkup data yang digunakan, dimana data yang diperoleh sebanyak 1008 tweet dari media sosial X, tidaklah mencakup seluruh pandangan warga masyarakat Indonesia terhadap fenomena golongan putih pada pemilihan umum 2024, melainkan penelitian yang dilakukan ini merepresentasikan 1008 tweet atau pandangan masyarakat Indonesia yang aktif menggunakan media sosial X dalam kurun waktu perode pengambilan data. Sehingga disarankan penelitian selajutnya untuk dapat menggunakan data dari berbagai media sosial, ataupun komentar yang ada dibanyak berita serta menggunakan data dari berbagai sumber online yang ada untuk memperoleh hasil yang lebih komprehensif, khususnya yang menggambarkan pandangan masyarakat Indonesia terhadap fenomena golongan putih pada pemilihan umum tahun 2024.

UCAPAN TERIMA KASIH

Ucapan terima kasih yang pertama kepada Tuhan Yesus Kristus yang tiada berkesudahan kasih setia-Nya dalam kehidupan penulis, sehingga penelitian ini dapat selesai dengan begitu baik. Mengucapkan terima kasih kepada orang tua terkasih, bapak Novri Saroinsong, ibu Agustiningsih Soeprijatno, dan saudari Zefanya Cantiq Saroinsong, yang senantiasa mendoakan, mendukung dan memberikan yang terbaik buat penulis. Mengucapkan terima kasih juga kepada semua pihak yang telah memberikan bantuan, dukungan, dan kontribusi dalam proses penyusunan artikel ini.

Penulis memahami bahwa penelitian ini masih belum sepenuhnya sempurna, oleh karena itu penulis dengan senang hati menerima masukan yang membangun demi kebaikan bersama. Semoga melalui penelitian yang dilakukan ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang analisis terhadap pandangan masyarakat menggunakan data dari media sosial. Dan akhir kata kiranya-Nya kasih karunia, damai sejahtera dari Tuhan Yesus Kristus senantiasa menyertai kita semua.

References

[1] C. Y. Tiouw and I. R. H. T. Tangkawarow, "Studi literatur penerapan WebGIS dan metode regresi linear dalam sistem monitoring dan prediksi kerawanan Pemilu," Edutik: Jurnal Pendidikan Teknologi Informasi dan Komunikasi, vol. 6, 2026.
[2] A. P. Wibowo, E. W. Wardhana, and T. H. Nurgiansah, "Pemilihan umum di Indonesia dalam perspektif Pancasila," Jurnal Kewarganegaraan, vol. 6, no. 2, 2022.
[3] Indraerawati and Rahmiati, "Golongan putih dalam pemilihan umum di Indonesia perspektif siyasah syar'iyyah," 2021.
[4] T. R. O. Tumimomor, I. R. H. T. Tangkawarow, and A. A. Kenap, "Analisis sentimen dan ujaran kebencian pemberitaan online tentang IKN menggunakan algoritma K-NN," The Indonesian Journal of Computer Science, vol. 14, no. 2, 2025.
[5] I. W. Suardi and I. R. H. T. Tangkawarow, "Perbandingan nilai akurasi analisa sentiment pada kata kunci Pemilu 2024," The Indonesian Journal of Computer Science, vol. 14, no. 2, 2025.
[6] S. N. Mawikere and I. R. H. T. Tangkawarouw, "Perbandingan akurasi SVM dan Naive Bayes pada analisis sentimen data berita online terhadap COKLIT Pemilu 2024," The Indonesian Journal of Computer Science, vol. 14, no. 2, 2025.
[7] E. Novianto, Suhirman, and D. Prasetyo, "Perbandingan metode klasifikasi Random Forest dan Support Vector Machine dalam memprediksi capaian studi mahasiswa," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 9, no. 4, pp. 1821–1833, 2024, doi: 10.29100/jipi.v9i4.5423.
[8] M. Kassa and I. R. H. T. Tangkawarow, "Studi literatur: Analisis penggunaan Support Vector Machine dalam klasifikasi teks berita distribusi logistik Pilkada 2024," Edutik: Jurnal Pendidikan Teknologi Informasi dan Komunikasi, vol. 6, 2026.
[9] A. A. Nurkhaliza and A. W. Wijayanto, "Perbandingan algoritma klasifikasi Support Vector Machine dan Random Forest pada prediksi status Indeks Mitigasi dan Kesiapsiagaan Bencana (IMKB) satuan kerja BPS di Indonesia tahun 2020," Jurnal Informatika Universitas Pamulang, vol. 7, no. 1, 2022, doi: 10.32493/informatika.v7i1.16117.
[10] D. A. Fitri and D. Damayanti, "Komparasi algoritma Random Forest Classifier dan Support Vector Machine untuk sentimen masyarakat terhadap pinjaman online di media sosial," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 9, no. 4, pp. 2018–2029, 2024, doi: 10.29100/jipi.v9i4.5608.
[11] I. Septiana and D. Alita, "Perbandingan Random Forest dan SVM dalam analisis sentimen quick count Pemilu 2024," Jurnal Informatika: Jurnal Pengembangan IT, vol. 9, no. 3, pp. 224–233, 2024, doi: 10.30591/jpit.v9i3.6640.
[12] P. K. Sari and R. R. Suryono, "Komparasi algoritma Support Vector Machine dan Random Forest untuk analisis sentimen metaverse," Mnemonic, vol. 7, no. 1, 2024.
[13] Z. C. Dwinnie, "Penerapan Machine Learning pada analisis sentimen Twitter sebelum dan sesudah debat calon presiden dan wakil presiden tahun 2024," Media Informatika Budidarma, vol. 8, no. 2, 2024, doi: 10.30865/mib.v8i2.7504.
[14] E. R. S. Moningkey, D. D. Harisondak, and K. Santa, "Support Vector Machine algorithm for classifying public satisfaction index," Academia Open, vol. 10, no. 2, 2025, doi: 10.21070/acopen.10.2025.12697.

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.14701

[15] R. R. Runtu, G. D. P. Maramis, and A. Hasibuan, "Implementasi algoritma rule-based classification pada aplikasi arsip web menggunakan metode RAD dan evaluasi ISO/IEC 25010," Jurnal Minfo Polgan, vol. 14, no. 2, pp. 2792–2798, 2025, doi: 10.33395/jmp.v14i2.15505.
[16] V. N. D. Abram, I. R. H. T. Tangkawarow, and A. A. Kenap, "Analisis kemiripan menggunakan metode BERT dan Jaccard pada proses bisnis Bawaslu RI," vol. 14, no. 3, 2025.
[17] A. Lahinda, I. R. H. T. Tangkawarow, and E. R. S. Moningkey, "Deteksi ujaran kebencian pada media sosial terkait Pemilu 2024 menggunakan algoritma Naive Bayes," Jurnal Minfo Polgan, vol. 14, no. 2, pp. 3140–3155, 2025, doi: 10.33395/jmp.v14i2.15665.
[18] A. A. Rosanti and F. Jannah, "Perbandingan metode Random Forest dan Support Vector Machine pada analisis sentimen Twitter terkait maskapai penerbangan," ResearchGate, 2023.
[19] F. B. Tambuwun, A. A. Kenap, and I. R. H. T. Tangkawarow, "Predicting Znergy athletes' swimming times using Random Forest based on lifestyle physiological exercise," Journal La Multiapp, vol. 7, no. 2, 2026.
[20] G. I. Ratumbuysang and I. R. H. T. Tangkawarow, "Studi literatur: Analisis topik otomatis isu Pemilu Indonesia menggunakan algoritma LDA dan NMF," Edutik: Jurnal Pendidikan Teknologi Informasi dan Komunikasi, vol. 6, 2026.
[21] P. W. L. Rompas, S. C. Kumajas, and Q. C. Kainde, "Penerapan algoritma binary search pada aplikasi E-Service Program Studi Teknik Informatika Universitas Negeri Manado berbasis web," Jurnal Minfo Polgan, vol. 14, 2025, doi: 10.33395/jmp.v14i2.15491.
[22] N. B. Sidauruk, N. Riza, and R. N. S. Fatonah, "Penggunaan metode SVM dan Random Forest untuk analisis sentimen ulasan pengguna terhadap KAI Access di Google Playstore," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 3, 2023.
[23] R. Mursyid and A. D. Indriyanti, "Perbandingan akurasi metode analisis sentimen untuk evaluasi opini pengguna pada platform media sosial (Studi kasus: Twitter)," Journal of Informatics and Computer Science, vol. 6, no. 2, 2024.
[24] H. Sumual, R. Rorimpandey, and F. Andries, "Pemanfaatan Youtube berbasis teknologi informasi dan komunikasi (TIK) dalam model project-based learning untuk meningkatkan pemahaman listening mahasiswa," Jurnal Pendidikan Teknologi Informasi dan Komunikasi, vol. 6, no. 1, 2026.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo