

# ANALISIS SENTIMEN TERHADAP PROGRAM KAMPANYE TABRAK PROF PADA MEDIA SOSIAL X DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Fauzan Maulana Lubis<sup>1</sup>, Muhammad Ikhsan<sup>2</sup>

<sup>1,2</sup> Jurusan Ilmu Komputeri Fakultas Sains dan Teknologi Universitas Islam Negeri Sumatera Utara  
Jl. Lap. Golf No. 120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara, 20353

<sup>1</sup>oanlubis82@gmail.com

<sup>2</sup>mhd.ikhsan@uinsu.ac.id

## Abstract

Pada Tahun 2024 Indonesia memasuki tahun demokrasi yaitu Pemilihan Presiden dan Wakil Presiden yang diadakan setiap 5 tahun sekali yang banyak menarik perhatian, Pasangan calon Presiden dan wakil Presiden Ganjar Pranowo dan Mahfud MD menghadirkan inovasi baru dalam berkampanye yaitu melalui program "Tabrak Prof" di mana mereka berinteraksi langsung dengan masyarakat untuk mendengarkan aspirasi dan keluh kesah secara terbuka. Tabrak Prof merupakan salah satu program kampanye yang dipelopori oleh Mahfud MD dalam menyampaikan visi dan misi serta program kerja yang digagasnya sebagai calon Wakil Presiden pada Pilpres 2024. Kampanye ini juga banyak mendapat simpati di berbagai kalangan, termasuk mahasiswa dan pekerja informal, serta memanfaatkan media sosial khususnya media sosial X untuk menyebarkan informasi dan berinteraksi dengan publik. Media sosial X dikenal sebagai pusat informasi cepat dan tepat, serta menjadi sarana efektif dan efisien untuk tempat promosi atau berkampanye. Penelitian ini di harapkan dapat bermanfaat dan membantu untuk melakukan riset atas opini masyarakat yang mengandung sentimen positif dan negatif melalui teknik pengolahan data tekstual. Pada penelitian ini, penulis menggunakan 944 data yang diambil dari media sosial X dengan metode *preprocessing* seperti *cleaning*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Penelitian ini memiliki signifikansi dalam konteks politik, khususnya dalam memahami opini publik terhadap program kampanye "Tabrak Prof" yang dijalankan oleh pasangan calon Ganjar Pranowo dan Mahfud MD. Hasil penelitian menunjukkan akurasi sebesar 81%, *precision* 81%, *recall* 90%, dan *F1-score* 85%, yang menunjukkan bahwa metode *Support Vector Machine* (SVM) efektif dalam menganalisis sentimen di media sosial.

Kata kunci : Tabrak Prof, Analisis Sentimen, Kampanye, *Support Vector Machine*

## I. PENDAHULUAN

Indonesia pada saat ini sedang memasuki tahun Demokrasi, yaitu Pemilihan Presiden dan Wakil Presiden yang diadakan setiap 5 tahun sekali, Komisi Pemilihan Umum (KPU) resmi merilis disetiap pasangan calon akan ada 5 kali Debat, 3 Debat untuk calon Presiden, 2 kali untuk calon Wakil Presiden, untuk menyampaikan Visi dan Misinya. selain debat KPU juga memberikan kesetiap pasangan calon berkampanye keseluruhan daerah yang ada di Indonesia.

Dalam Atmosfer Kampanye tahun ini, masing-masing pasangan calon punya trobosan yang menarik perhatian untuk meyakinkan kepada masyarakat dalam pemilihan pemilu nantinya. pada kampanye di tahun 2019 kemarin, setiap pasangan calon hanya menyampaikan Visi dan Misi diatas panggung saja tidak ada sesuatu yang baru. Tetapi ditahun ini ada hal yang menarik dari pasangan calon Ganjar Pranowo dan

Mahfud MD, Mahfud merubah cara berkampanye disetiap daerah yang mereka datang.

Tabrak prof, dimana cara berkampanye mereka langsung terjun ke daerah dengan membuka ruang untuk mendengarkan keluh kesah Masyarakat dan Aspirasi Masyarakat selama beberapa tahun ini. Dalam forum Tabrak Prof, semua pertanyaan bisa diajukan oleh masyarakat yang mencakup sesuai dengan Visi dan Misi. Selain interaksi langsung dengan masyarakat, program ini dapat juga dikhususkan untuk berbagai kalangan anak muda seperti Mahasiswa, bahkan ojol sekalipun. Pada saat ini, media sosial X dapat dijadikan sebuah indikator yang baik untuk memberikan pengaruh di dalam sebuah penelitian. isi *tweet* di X dijadikan objek penelitian. Namun tidak semua *tweet* dijadikan objek penelitian. Biasanya *tweet* hanya topik-topik tertentu yang dijadikan objek penelitian, seperti *tweet* yang berhubungan dengan dunia politik, budaya, pemasaran, dan sebagainya. tak sedikit Perusahaan dan tokoh-

tokoh berpengaruh yang mempunyai akun X untuk menyapa dan berinteraksi dengan pengguna lain di media sosial X [1].

X digunakan sebagai analisis sentimen pada studi kasus berbagai bidang karena banyak dijadikan sebagai sumber data. Selain itu, X menjadi pusat trending di Indonesia. X merupakan media sosial yang mempunyai penyebaran informasi paling cepat dan tepat dari media sosial lainnya. X mempunyai karakteristik yaitu dapat mengirim opini tanpa batas, mencari berita terkini, membagikan tweet pengguna lain dan memberi komentar. Adanya karakteristik X tersebut, para peneliti semakin mudah dalam menganalisis karena pengguna secara bebas dapat mengungkapkan pendapatnya, sehingga X menjadi sarana terbaik untuk melakukan Analisis Sentimen masyarakat [2].

Analisis Sentimen adalah proses mengidentifikasi dan mengelompokkan opini yang masih berbentuk teks ke dalam sentimen positif atau negatif. Dalam 3 hal ini, penerapan Analisis Sentimen dapat berupa menganalisa pendapat, sentimen, evaluasi, emosi, penilaian, atau sikap pada suatu produk, tokoh, organisasi, layanan, isu, atau peristiwa yang terjadi di masyarakatnya sendiri. Selain itu, Analisis Sentimen memiliki keterkaitan erat dengan masyarakat, karena sumber informasi yang digunakan berasal dari media sosial, yang penggunaannya adalah masyarakat itu sendiri [3].

Analisis Sentimen, atau dikenal sebagai *Opinion Mining*, adalah proses otomatis untuk memahami, mengekstrak, dan mengolah data teks guna mengidentifikasi sentimen dalam suatu opini. Karena pengaruh dan manfaatnya yang besar, penelitian serta pengembangan aplikasi berbasis analisis sentimen terus berkembang pesat [4].

Penelitian ini menerapkan algoritma *Support Vector Machine* (SVM). Metode SVM merupakan teknik untuk meningkatkan margin pemisahan antara dua kelas yang berbeda. Menurut penelitian sebelumnya, Metode *Support Vector Machine* (SVM) Para peneliti banyak yang menggunakan Metode tersebut pada X untuk melakukan Analisis Sentimen [5].

Metode *Support Vector Machine* (SVM) dipilih dalam penelitian ini karena kemampuannya dalam menangani data yang memiliki dimensi tinggi dan non-linear, serta performanya yang telah terbukti dalam berbagai penelitian terkait analisis sentimen. Menurut Alfredo Gormantara (2020), SVM menunjukkan akurasi yang lebih tinggi dibandingkan metode *Naive Bayes* dalam analisis sentimen di media sosial. Selain itu, SVM juga lebih efektif dalam menangani data yang tidak seimbang, seperti dalam kasus *dataset* yang digunakan dalam penelitian ini [6]. Referensi terkini seperti Hashri hayati et al., (2021) juga mendukung penggunaan SVM untuk analisis sentimen di media sosial karena kemampuannya dalam mengklasifikasikan data dengan margin yang optimal [7].

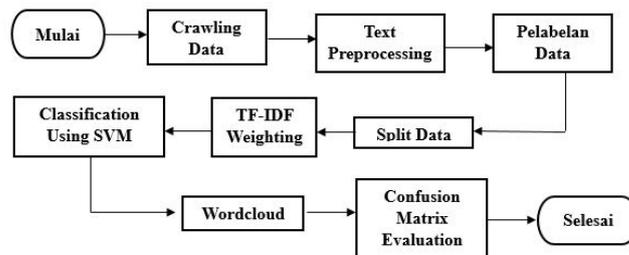
Adanya program kampanye Tabrak Pof yang mengundang banyak kontroversi di media sosial khususnya X baik positif maupun negatif menjadi suatu hal yang harus diketahui oleh publik. Dengan adanya penelitian ini dapat mengetahui pendapat publik di media sosial X mengenai Tabrak prof apakah cenderung Positif atau Negatif. Oleh karena itu, penulis mengangkat tema penelitian dalam bentuk skripsi dengan judul "Analisis Sentimen Terhadap Program Kampanye Tabrak Prof

Pada Media Sosial X dengan menggunakan Metode *Support Vector Machine*" [8].

## II. METODOLOGI PENELITIAN

Penelitian ini menggunakan Metodologi Kualitatif yang Penelitiannya berfokus pada elemen manusia, objek dan institusi, serta hubungan atau interaksi di antara elemen-elemen tersebut dalam upaya memahami suatu peristiwa, perilaku atau fenomena. dalam penelitian ini menjelaskan secara rinci dan lengkap tentang fenomena yang terjadi di Media Sosial X, tentang analisis sentimen yang mengekstrak sentimen *tweet* tabrak prof dengan menerapkan Algoritma *Support Vector Machine* (SVM) [9].

Proses *crawling data* dilakukan menggunakan *library Tweepy*, yang memungkinkan pengambilan data dari media sosial X secara otomatis. Data yang diambil difilter berdasarkan *hashtag* "Tabrak Prof" dan "Mahfud MD" untuk memastikan relevansi dengan topik penelitian. Untuk memastikan validitas data, peneliti melakukan proses *cleaning* dan *preprocessing* untuk menghilangkan noise seperti spam, duplikat, atau konten yang tidak relevan. Selain itu, *dataset* yang digunakan telah melalui proses validasi dengan membandingkan distribusi sentimen positif dan negatif untuk memastikan bahwa data tidak bias. Dalam implementasi SVM, *kernel linear* dipilih karena data yang digunakan memiliki karakteristik linear dan kernel ini telah terbukti efektif dalam menangani dataset dengan dimensi yang relatif kecil [10].



Gambar 1. Tahapan Penelitian

### A. Crawling Data

Proses pelaksanaan penelitian ini membutuhkan data yang *real* untuk dilakukan pemrosesan suatu data, peneliti mengambil data dari platform media sosial X menggunakan *hashtag* Tabrak prof.

### B. Text Preprocessing

Pada tahap berikutnya, setelah data diperoleh dari media sosial X, dilakukan proses *Preprocessing* melalui lima langkah, yaitu *Cleaning*, *Tokenizing*, *Normalisasi*, *Stopword Removal*, *Stemming*, dan *Detokenized Text*.

### C. Pelabelan data

Pelabelan data pada penelitian ini akan menggunakan pelabelan data otomatis dengan *Inset Lexicon* dimana pelabelan akan menjadi 2 yaitu positif dan negatif. Berikut ini adalah hasil proses pelabelan data.

### D. Split data

Pada tahap ini, *dataset* akan dibagi menjadi dua yaitu, data latih dan data uji. Dimana pada penelitian ini perbandingan antara data latih dan data uji adalah 8:2. dengan *dataset* yang berjumlah 755 data latih dan 189 data uji.

#### E. TF-IDF Weighting

Tahap pembobotan adalah proses pemberian nilai pada setiap kata menggunakan perhitungan *Term Frequency–Inverse Document Frequency* (TF-IDF). Tujuan dari pembobotan ini adalah untuk menentukan skor pada setiap term, yang nantinya akan digunakan sebagai input dalam proses klasifikasi.

#### F. Classification using SVM

Pada tahap berikutnya, setelah data selesai melalui lima langkah *preprocessing*, proses klasifikasi dilakukan menggunakan skenario pemodelan dengan *Linear Kernel SVM*. Teknik *Support Vector Machine* (SVM) bertujuan untuk menemukan fungsi pemisah optimal di antara berbagai fungsi yang tersedia guna membedakan dua jenis objek. Dalam proses klasifikasi SVM, diperlukan bobot untuk setiap kata dalam data pelatihan. Nilai bobot dari masing-masing term diperoleh melalui proses pembobotan TF-IDF, yang kemudian digunakan dalam pembuatan model dengan algoritma SVM. .

#### G. Wordcloud

*Wordcloud* adalah representasi visual dari kumpulan kata yang sering muncul dalam *tweet*. Fungsinya adalah untuk mempermudah dalam memahami topik yang sering dibahas oleh pengguna di suatu platform, khususnya media sosial.

#### H. Confusion Matrix Evaluation

*Confusion Matrix Evaluation* dilakukan untuk menilai kinerja model dalam proses klasifikasi. Terdapat beberapa metode yang umum digunakan untuk mengukur performa, di antaranya adalah perhitungan *recall*, *precision*, dan *F1-Score*. Metode evaluasi yang diterapkan, yaitu *Confusion Matrix*, yang menghasilkan nilai akurasi, *precision*, *recall*, dan *F1-Score* berdasarkan perbandingan antara hasil prediksi dan data aktual.

### III. HASIL DAN PEMBAHASAN

Dalam penelitian ini, penulis mengumpulkan data dari media sosial X (Twitter). Komentar diperoleh melalui proses *Crawling* dan disimpan dalam file berformat .csv. Total data yang dikumpulkan berjumlah 944 komentar yang membahas sentimen terkait "Tabrak Prof" atau opini pengguna aplikasi X mengenai topik tersebut. Selanjutnya, komentar diklasifikasikan dengan label positif dan negatif menggunakan kamus *Inset Lexicon* serta Algoritma *Support Vector Machine* untuk mengukur akurasi, *precision*, *recall*, dan *F1-Score*. [11].

#### A. Pengumpulan data

Data komentar diperoleh dari platform X pada periode Juni 2024 dengan menggunakan Google Colab sebagai alat aksesnya. Sumber data penelitian ini berupa ulasan yang dikumpulkan dari komentar para pengguna aplikasi X. Contoh ulasan yang digunakan ditunjukkan pada Tabel dibawah.

**Tabel 1.** Komentar Pengguna Aplikasi X Pada Sentimen Tabrak Prof

Waktu <i>tweet</i>	<i>tweet</i>	<i>Username</i>
Thu Feb 08 13:19:10	@mohmahfudmd Setuju prof.. Pokoke gaspol GAMA! #tabrakprof	@WidodoNPRJKT

#### B. Cleaning

Tahap *Cleaning* dilakukan untuk menyaring dan memurnikan data komentar yang telah dikumpulkan. Proses ini mencakup penghilangan *username* X, penghapusan tautan, eliminasi tanda baca, standardisasi teks menjadi huruf kecil, serta pembersihan emotikon dan elemen-elemen *non*-tekstual lainnya. Tujuannya adalah menghasilkan data teks yang terstandarisasi dan bebas dari unsur-unsur yang tidak diperlukan untuk analisis.

**Tabel 2.** Data sebelum dan sesudah melewati tahap *Text clean*

<i>tweet</i>	<i>cleaning</i>
@mohmahfudmd Setuju prof.. Pokoke gaspol GAMA! #tabrakprof	setuju prof pokoke gaspol gama tabrakprof

#### C. Tokenizing

Tokenisasi adalah proses dalam pemrosesan bahasa alami *Natural Language Processing* (NLP) yang memecah teks menjadi unit-unit kecil yang disebut token. Token bisa berupa kata, frasa, atau karakter individu, tergantung dari kebutuhan analisisnya. Proses ini memudahkan pemrosesan teks lebih lanjut, seperti analisis sentimen, pemrosesan kalimat, atau penghitungan frekuensi kata.

**Tabel 3.** Data sebelum dan sesudah melewati tahap *Tokenizing*

<i>Text clean</i>	<i>Tokenizing</i>
setuju prof pokoke gaspol gama tabrakprof	"['setuju', 'prof', 'pokoke', 'gaspol', 'gama', 'tabrakprof']"

#### D. Normalisasi

Normalisasi merupakan proses transformasi kata-kata *non*-baku seperti mengubah "yg" menjadi bentuk bakunya "yang", atau "ga" menjadi "tidak", yang bertujuan untuk menstandarisasi teks sesuai dengan kaidah bahasa Indonesia yang benar berdasarkan KBBI.

#### E. Stopword Removal

Proses *stopword removal* diterapkan untuk mengeliminasi kata-kata yang tidak memiliki kontribusi signifikan dalam analisis sentimen. Sistem akan membandingkan setiap kata dalam *tweet* dengan daftar *stopword* yang tersimpan dalam database, kemudian menghapus kata-kata yang teridentifikasi sama. Eliminasi ini dilakukan karena kata-kata stop tidak mengandung makna spesifik atau esensial, sehingga dapat dianggap sebagai 'derau' yang mengganggu dalam pemrosesan teks."

**Tabel 4.** Data sebelum dan sesudah melewati tahap *Stopword Removal*

<i>Tokenizing</i>	<i>Stopword Removal</i>
"['setuju', 'prof', 'poko', 'gaspol', 'gama', 'tabrakprof']"	"['setuju', 'prof', 'poko', 'gaspol', 'gama', 'tabrakprof']"

**F. Stemming**

Proses *Stemming* diterapkan untuk mengurangi dimensi data dan meningkatkan efisiensi dalam analisis teks serta pencarian informasi. Teknik ini mengubah kata-kata ke bentuk dasar atau akar katanya, sehingga kata-kata yang berbeda namun memiliki makna serupa dapat diperlakukan sebagai satu kesatuan.

**Tabel 5.** Data sebelum dan sesudah melewati tahap *Stemming*

<i>Stopword Removal</i>	<i>Stemming</i>
"['setuju', 'prof', 'poko', 'gaspol', 'gama', 'tabrakprof']"	"['tju', 'prof', 'poko', 'gaspol', 'gama', 'tabrakprof']"

**G. Detokenized Text**

*Detokenized* adalah proses yang menghapus tanda “[’]” dalam tahap *stemming*, sehingga hasilnya dapat dijadikan acuan sebagai output akhir dari proses preprocessing yang telah dilakukan.

**Tabel 6.** Data sebelum dan sesudah melewati tahap *detokenized text*

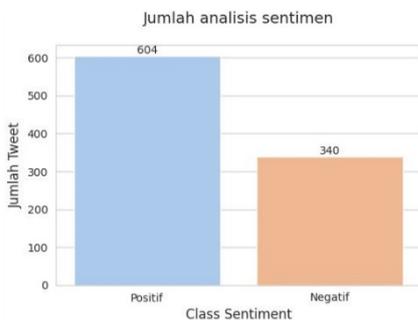
<i>Stemming</i>	<i>Detokenized Text</i>
"['tju', 'prof', 'poko', 'gaspol', 'gama', 'tabrakprof']"	tuju prof poko gaspol gama tabrakprof

**H. Pelabelan data**

Tahap selanjutnya setelah *text preprocessing* yaitu pelabelan data. Pelabelan data pada penelitian ini akan menggunakan pelabelan data otomatis dengan kamus *Inset Lexicon*.

**Tabel 7.** Pelabelan *dataset*

No	<i>completepraproses</i>	label
1	tuju prof poko gaspol gama tabrakprof	Positif



**Gambar 2.** Pelabelan

**I. Split data**

Setelah melalui tahapan pelabelan, data selanjutnya akan dibagi atau bisa disebut juga *splitting* data. Pada tahapan ini, data akan dibagi menjadi data uji dan data latih. Dimana data latih sebesar 80% dan data uji sebesar 20%. Berikut contoh sampel data latih serta data uji.

**Tabel 8.** Tabel data latih

sama seruduk tabrakprof
semoga prof mahfud md segera sembuh sehat sedia
mulia lahir salah tokoh bangsa integritas amin
tabrakprof rusak rusak

**Tabel 9.** Tabel data uji

curiga mahfud md orang kirim gagal in anies imin ya
gue mahfud md

**J. Pembobotan Kata Menggunakan TF-IDF**

Setelah proses pelabelan dan *preprocessing*, langkah berikutnya adalah tahap pembobotan menggunakan metode TF-IDF. Pada tahap ini, digunakan teknik untuk menghitung bobot setiap kata (term) dalam dokumen, di mana setiap kata dikalikan dengan nilai IDF. Perhitungan nilai TF dan DF dari lima data latih serta data uji disajikan dalam bagian ini.

**Tabel 10.** Nilai DF pada data latih dan data uji

Kata (term)	D1	D2	D3	D4	D5	Total DF
mahfud md	0	1	0	1	1	3
tabrakprof	1	0	1	0	0	2
...	...	...	...	...	...	...
imin	0	0	0	1	0	1

Setelah nilai TF di dapatkan, selanjutnya mencari nilai IDF. Berikut rumus mencari IDF pada setiap kata: Diketahui D = 5 dan DF = 3  
Penyelesaian:

$$IDF = \log \frac{5+1}{3+1} + 1$$

$$IDF = \log \frac{6}{3} + 1 = 1.1761$$

Hasil nilai IDF tertera berikut:

**Tabel 11.** Nilai IDF pada data latih dan data uji

Kata	DF	IDF
mahfud md	3	1.1761
tabrakprof	2	1.3010
...	...	...
imin	1	1.4771

Setelah diketahui nilai TF dan IDF di peroleh selanjutnya adalah menghitung nilai TF-IDF nya  
 $W = TF \times IDF$

**Tabel 12.** Nilai TF-IDF pada data latih dan data uji

Kata (term)	D1	D2	D3	D4	D5
mahfud md	0	1.4771	0	1.4771	1.4771
tabrakprof	1.4771	0	1.4771	0	0
...	...	...	...	...	...
imin	0	0	0	1.4771	0

Setelah nilai TF-IDF di peroleh maka langkah selanjutnya tahap normalisasi, untuk menyamakan *interval* pada setiap data, berikut persamaan yang di gunakan untuk normalisasi data adalah:

$$W_{norm}(\text{mahfud md}) = \frac{TF-IDF(\text{mahfud md})}{\sqrt{\sum_i (TF-IDF(t_i, d))^2}}$$

Berikut adalah hasil perhitungan normalisasi data yang di lakukan. Hasil normalisasi data tertera pada tabel berikut ini:

**Tabel 13.** Normalisasi pada data latih dan data uji

No	D1	D2	D3	D4	D5
1	0	0.0591	0	0.0591	0.0591
2	0.0591	0	0.0591	0	0
...	...	...	...	...	...
24	0	0	0	0.0591	0

### K. Klasifikasi Support Vector Machine

Pasca proses pembersihan dan strukturisasi data, tahap berikutnya adalah penerapan klasifikasi menggunakan Algoritma Support Vector Machine. Klasifikasi diawali dengan pembagian *dataset* menjadi dua kelompok yaitu data *training* dan data *testing*. Set data *training* dimanfaatkan untuk mempelajari pola dan karakteristik pembeda antara kategori positif dan negatif, sementara set data *testing* berfungsi untuk mengukur tingkat keberhasilan model dalam melakukan prediksi klasifikasi secara tepat. Dalam implementasi algoritma ini, *kernel linear* dipilih sebagai fungsi transformasi karena sifat data yang diproses merupakan *data linear* [12].

```
# Mengubah teks ke bentuk vektor
vectorizer = CountVectorizer()
X_vectorized = vectorizer.fit_transform(X)

# Membagi data ke dalam training dan testing set
X_train, X_test, y_train, y_test = train_test_split(X_vectorized, y, test_size=0.2, random_state=42)

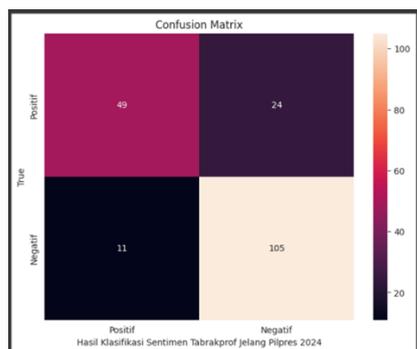
# Melatih model SVM
svm = SVC(kernel = "linear")
svm_model = svm.fit(X_train, y_train)

# Memprediksi data test
y_pred = svm_model.predict(X_test)
```

**Gambar 3.** Klasifikasi SVM

### L. Confusion Matrix Evaluation

Setelah proses analisis data selesai dilakukan, diperoleh hasil berupa label pada data uji yang diklasifikasikan menggunakan algoritma Support Vector Machine (SVM). Evaluasi model tidak hanya didasarkan pada nilai akurasi, tetapi juga menggunakan *classification report* untuk menilai performa klasifikasi yang telah diterapkan. Selain itu, hasil dari *Confusion Matrix* ditampilkan dalam tabel berikut. [13].



**Gambar 4.** Visualisasi Confusion Matrix

**Tabel 14.** Confusion Matrix

Aktual	Prediksi	Prediksi
Positif	49	24
Negatif	11	105

Dari tabel diatas, maka dapat dihitung nilai Akurasi, *Precision*, *Recall* serta *F1-score* dengan menggunakan persamaan berikut:

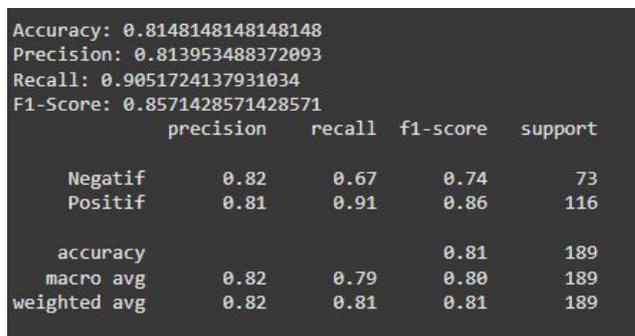
$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{49+105}{49+105+24+11} \times 100\% = 81\%$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{49}{49+24} \times 100\% = 81\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{49}{49+11} \times 100\% = 90\%$$

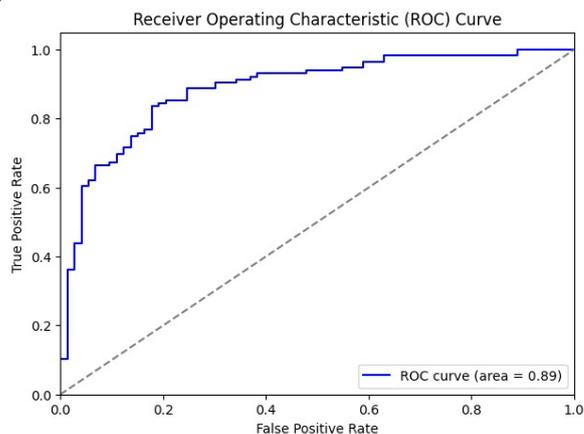
$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.813 \times 0.90}{0.813 + 0.90} \times 100\% = 85\%$$

Untuk menilai Akurasi, *Precision*, *Recall*, dan *F1-score* secara keseluruhan, dapat digunakan *classification report*. Berikut adalah *classification report* dari hasil analisis sentimen menggunakan Algoritma Support Vector Machine [14].



**Gambar 5.** Classification Analisis Tabrak prof

Hasil penelitian menunjukkan bahwa nilai *recall* (90%) lebih tinggi dibandingkan *precision* (81%), yang mengindikasikan bahwa model cenderung lebih baik dalam mengidentifikasi sentimen positif daripada negatif. Untuk memastikan tidak terjadi *overfitting*, peneliti melakukan validasi silang (*cross-validation*) dan mengevaluasi model menggunakan *confusion matrix*. Selain itu, *ROC curve* ditambahkan untuk memberikan gambaran yang lebih komprehensif tentang performa model [15].



**Gambar 6.** Kurva Receiver Operating Characteristic (ROC)

Kurva *Receiver Operating Characteristic* (ROC) yang dihasilkan menunjukkan bahwa model *Support Vector Machine* (SVM) memiliki *Area Under the Curve* (AUC) sebesar 0,89, yang menandakan bahwa model memiliki kemampuan klasifikasi yang sangat baik. ROC curve menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR), di mana semakin tinggi nilai AUC, semakin baik model dalam membedakan antara kelas positif dan negatif. Kurva yang berada jauh di atas garis diagonal menunjukkan bahwa model dapat mengidentifikasi kelas positif dengan tingkat keakuratan yang tinggi, meskipun masih terdapat kemungkinan kesalahan klasifikasi pada beberapa kasus [16].

#### IV. KESIMPULAN

Penelitian ini menganalisis opini masyarakat di Media Sosial X menggunakan metode *Support Vector Machine* (SVM) dengan membangun model pembelajaran sistem sehingga mampu melakukan analisis sentimen dalam waktu komputasi yang efisien. Pelabelan data dilakukan secara otomatis menggunakan *Inset Lexicon* pada 944 data yang telah melalui tahap *preprocessing text*. Data kemudian diklasifikasikan ke dalam kelas positif dan negatif. Pada analisis ini, data dibagi dengan perbandingan 8:2 antara data latih dan data uji. Dari total 944 data setelah *preprocessing*, 755 digunakan sebagai data latih dan 189 sebagai data uji. Tingkat akurasi yang dihasilkan oleh metode SVM dalam klasifikasi sentimen (opini masyarakat) menunjukkan hasil yang baik. Penelitian ini menggunakan dataset yang diperoleh melalui teknik *crawling* sebanyak 944 *tweet* dengan kata kunci "tabrak prof", "kampanye" dan "Mahfud MD". Berdasarkan dataset tersebut, diperoleh hasil akurasi sebesar 81%, precision 81%, recall 90%, dan f1-score 85%.

Penelitian ini juga menyimpulkan bahwa metode *Support Vector Machine* (SVM) efektif dalam menganalisis sentimen terhadap program kampanye "Tabrak Prof" di media sosial X. Namun, untuk penelitian lanjutan, disarankan untuk menggunakan metode lain seperti *Deep Learning* atau menambah jumlah dataset untuk meningkatkan akurasi model. Implikasi praktis dari penelitian ini adalah memberikan wawasan bagi tim kampanye politik dalam merancang strategi yang lebih efektif berdasarkan analisis sentimen publik. Implikasi praktis dari penelitian ini adalah memberikan wawasan bagi tim kampanye politik dalam merancang strategi yang lebih efektif berdasarkan analisis sentimen publik. Selain itu, penelitian ini juga dapat diterapkan dalam konteks lain, seperti analisis sentimen terhadap kebijakan pemerintah atau produk komersial, untuk memahami opini publik dan meningkatkan pengambilan keputusan.

Penelitian ini memiliki beberapa keterbatasan, seperti kemungkinan *overfitting* akibat distribusi data yang tidak seimbang dan keterbatasan metode SVM dalam menangani *dataset* yang sangat besar. Untuk penelitian lanjutan, disarankan untuk menggunakan teknik *oversampling* atau *undersampling* untuk mengatasi ketidakseimbangan data, serta mengeksplorasi metode lain seperti *Deep Learning* untuk meningkatkan performa model.

#### V. SARAN

Dalam penelitian selanjutnya, untuk meningkatkan nilai Akurasi, disarankan melakukan pengujian dengan berbagai model pelatihan dan metode yang berbeda serta menambah jumlah *dataset* yang digunakan.

#### REFERENSI

- [1] Af'idah, D. I., Dairoh, D., Handayani, S. F., & Pratiwi, R. W. (2021). Pengaruh Parameter Word2Vec terhadap Performa Deep Learning pada Klasifikasi Sentimen. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 156–161. <https://doi.org/10.30591/jpit.v6i3>.
- [2] Ananda, F. D., & Pristyanto, Y. (2021). Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(2), 407–416. <https://doi.org/10.30812/matrik.v20i2.1130>.
- [3] Ikhsan, M., Armansyah, A., & Tamba, A. A. (2022). Implementasi Jaringan Syaraf Tiruan Backpropagation Pada Klasifikasi Grade Teh Hitam. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 387. <https://doi.org/10.30865/json.v4i2.5312>.
- [4] Darwis, D., Pratiwi, E. S., & Pasaribu, A. F. O. (2020). Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Eductic - Scientific Journal of Informatics Education*, 7(1), 1–11. <https://doi.org/10.21107/edutic.v7i1.8779/>
- [5] Fathonah, F., & Herliana, A. (2021). Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes. *Jurnal Sains Dan Informatika*, 7(2), 155–164. <https://doi.org/10.34128/jsi.v7i2.331/>
- [6] Hayati, H., & Alifi, M. R. (2021). ANALISIS SENTIMEN PADA TWEET TERKAIT VAKSIN COVID-19 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE. 7(September), 110–119.
- [7] Gormantara, A. (2020). Analisis Sentimen Terhadap New Normal Era di Indonesia pada Twitter Analisis Sentimen Terhadap New Normal Era di Indonesia pada Twitter Menggunakan Metode Support Vector Machine. July.
- [8] Hasri, C. F., & Alita, D. (2022). PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN TERHADAP DAMPAK VIRUS CORONA DI TWITTER. 3(2), 145–160.
- [9] Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., & Samudi, S. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 293. <https://doi.org/10.24114/cess.v5i2.18186>.
- [10] Hermawansyah, A., & Pratama, A. R. (2021). Analisis Profil dan Karakteristik Pengguna Media Sosial di Indonesia dengan Metode EFA dan MCA Analysis of Profiles and Characteristics of Social Media Users in

- Indonesia using EFA and MCA Methods. Februari, 20(1), 69–82. <https://dspace.uui.ac.id/handle/123456789/42383>.
- [11] Hokijuliandy, E., Napitupulu, H., & Firdaniza, F. (2023). Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square. *SisInfo: Jurnal Sistem Informasi Dan Informatika*, 5(2), 40–49. <https://doi.org/10.37278/sisinfo.v5i2.670>
- [12] Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 31. <https://doi.org/10.33365/jdmsi.v2i1.1021>.
- [13] Natasuwarna, A. P. (2020). Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring. *Techno.Com*, 19(4), 437–448. <https://doi.org/10.33633/tc.v19i4.4044>.
- [14] Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711.
- [15] Riadi, I., Fadlil, A., & Murni, M. (2023). Identifying Hate Speech in Tweets with Sentiment Analysis on Indonesian Twitter Utilizing Support Vector Machine Algorithm. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, 9(2), 179–191. <https://journals.ums.ac.id/index.php/khif/article/view/22470>