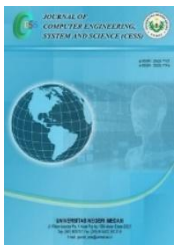
CESS (Journal of Computer Engineering, System and Science)
Vol. 10, No. 2, Juli 2025, pp.399-409

CESS
(Journal of Computer Engineering, System and Science)
Available online: https://jurnal.unimed.ac.id/2012/index.php/cess
ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Analisis Sentimen Terkait Kasus Korupsi Timah Pada Aplikasi X Menggunakan Algoritma Support Vector Machine

Sentiment Analysis Regarding the Tin Corruption Case on the X Platform Using the Support Vector Machine Algorithm

Suci Addina Ramadani Hsb[1*], Mhd. Furqon[2]

[1,2]Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara Medan, Indonesia
Email: [1]suciaddina785@gmail.com, [2]mfurqan@uinsu.ac.id
**Corresponding Author*

A B S T R A K

Kasus dugaan korupsi yang melibatkan PT Timah Tbk dengan estimasi kerugian negara sebesar Rp 271 triliun telah memicu respons luas dari masyarakat, khususnya melalui media sosial X (dahulu Twitter). Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kasus tersebut menggunakan algoritma *Support Vector Machine* (SVM). Data diperoleh dari platform X menggunakan kata kunci tertentu, menghasilkan 107 tweet yang kemudian melalui proses pra-pemrosesan dan pelabelan otomatis dengan metode berbasis leksikon (lexicon-based). Proses klasifikasi sentimen dilakukan dengan algoritma SVM, sementara ekstraksi fitur dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 98%, dengan *precision* 98% dan *recall* 100% untuk sentimen negatif, serta *precision* 100% dan *recall* 60% untuk sentimen positif. Berdasarkan *confusion matrix*, sebesar 95,33% data diklasifikasikan sebagai negatif secara benar (*True Negative*), 2,80% sebagai positif secara benar (*True Positive*), dan 1,87% salah diklasifikasikan (*False Negative*), tanpa terdapat kesalahan klasifikasi positif palsu (*False Positive*). Temuan ini menunjukkan dominasi opini negatif terhadap kasus tersebut dan menegaskan efektivitas SVM dalam analisis sentimen publik. Penelitian ini diharapkan menjadi acuan dalam memahami persepsi masyarakat dan pengambilan keputusan berbasis data terhadap isu sosial-politik di Indonesia.

Kata Kunci: *Analisis Sentimen; Support Vector Machine; Kasus Korupsi; Media Sosial; Berbasis Leksikon; TF-IDF; PT Timah; Opini Publik.*

A B S T R A C T

The alleged corruption case involving PT Timah Tbk, with estimated state losses of IDR 271 trillion, has triggered a widespread public reaction, particularly on the social media platform

This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license

X (formerly Twitter). This study aims to analyze public sentiment toward the case using the Support Vector Machine (SVM) algorithm. A total of 107 tweets were collected using specific keywords and underwent a series of preprocessing steps followed by automatic sentiment labelling using a lexicon-based method. Sentiment classification was then carried out using the SVM algorithm, with feature extraction performed through the Term Frequency-Inverse Document Frequency (TF-IDF) technique. Evaluation results showed the model achieved an accuracy of 98%, with a precision of 98% and recall of 100% for negative sentiment, and a precision of 100% and recall of 60% for positive sentiment. Based on the confusion matrix, 95.33% of the data were correctly classified as negative (True Negative), 2.80% as positive (True Positive), and 1.87% were misclassified (False Negative), with no false positives recorded. These findings highlight a dominant negative public perception of the corruption issue and demonstrate the effectiveness of SVM in sentiment analysis on social media. This research contributes to understanding public opinion and supports data-driven decision-making regarding socio-political issues in Indonesia.

Keywords: *Sentiment Analysis; Support Vector Machine; Corruption Case; Social Media; Lexicon-Based; TF-IDF; PT Timah; Public Opinion.*

1. PENDAHULUAN

Kasus dugaan korupsi di PT. Timah Tbk yang menimbulkan kerugian negara hingga mencapai Rp 271 triliun telah menimbulkan kegemparan publik di Indonesia. Skandal ini tidak hanya mengguncang sektor industri pertambangan, tetapi juga menggoyahkan kepercayaan masyarakat terhadap tata kelola sumber daya alam dan integritas pejabat publik. Salah satu wujud nyata reaksi publik terhadap isu ini terlihat dari ribuan opini yang tersebar di media sosial, terutama platform X (sebelumnya Twitter), yang menjadi ruang ekspresi dan diskusi digital masyarakat terhadap isu-isu sosial-politik.

Perkembangan pesat media sosial sebagai media penyebaran opini publik mendorong pentingnya pemanfaatan teknologi analisis sentimen untuk memahami persepsi masyarakat terhadap suatu isu[1]. Analisis sentimen memanfaatkan teknik natural language processing (NLP) dan machine learning untuk mengklasifikasikan opini dalam bentuk positif, negatif, atau netral. Untuk tugas klasifikasi semacam ini, algoritma Support Vector Machine (SVM) menjadi salah satu metode yang paling banyak digunakan karena kehandalan nya dalam menangani data berdimensi tinggi dan pemisahan kelas yang kompleks[2][3].

Penelitian-penelitian sebelumnya menunjukkan performa tinggi SVM dalam klasifikasi sentimen pada berbagai konteks[4] mencatat akurasi sebesar 77% dalam menganalisis sentimen pengguna aplikasi BSI Mobile [5]. menunjukkan hasil yang sangat baik dengan akurasi mencapai 98% dalam analisis sentimen masyarakat terhadap pandemic COVID-19. menerapkan SVM untuk mengklasifikasikan opini publik terhadap calon presiden RI dan mendapatkan akurasi rata-rata sebesar 78% [6] Menekankan efektivitas SVM dalam mengklasifikasikan opini berbasis fitur teks mendalam, sementara [7] menunjukkan bahwa margin maksimal SVM dapat mengoptimalkan pemisahan kelas dalam data sosial yang kompleks.

Selain itu, algoritma ini juga telah digunakan dalam konteks pelayanan publik dan transportasi online[8]. Menunjukkan bahwa SVM mampu mengklasifikasikan sentimen pelanggan dengan nilai precision dan recall yang tinggi, mencapai lebih dari 90%. Penelitian

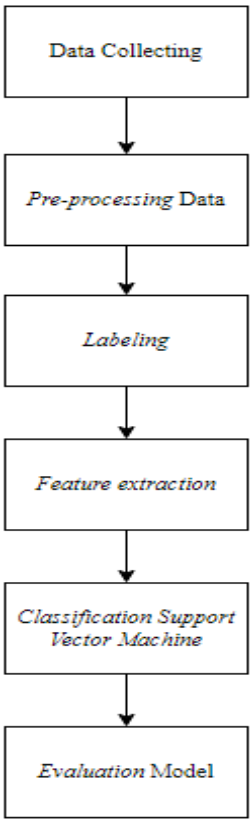
lain membuktikan efektivitas SVM dalam mengklasifikasikan sentimen publik terhadap debat calon wakil presiden Indonesia tahun 2024[1].

Dengan mempertimbangkan latar belakang tersebut, penelitian ini bertujuan untuk mengaplikasikan algoritma Support Vector Machine dalam klasifikasi sentimen masyarakat terhadap kasus korupsi timah, berdasarkan data dari media sosial X. Penelitian ini diharapkan mampu memberikan gambaran empiris tentang persepsi publik serta memperkaya literatur ilmiah dalam penerapan machine learning untuk analisis isu sosial di Indonesia.

Namun demikian, sebagian besar penelitian sebelumnya masih berfokus pada isu-isu seperti layanan publik, pandemic, atau politik elektoral, dan belum banyak yang mengkaji isu korupsi di sektor pertambangan yang berdampak besar seperti kasus PT Timah Tbk. Selain itu, belum ada studi yang secara khusus memanfaatkan media sosial X sebagai sumber data utama dalam memahami dinamika opini publik terhadap tata kelola sumber daya alam. Kondisi ini menunjukkan adanya celah penelitian yang perlu diisi, yaitu penerapan analisis sentimen berbasis SVM dalam konteks skandal sosial-ekonomi berskala besar yang mencerminkan ketidakpercayaan publik secara digital. Oleh karena itu, penelitian ini berupaya menutup celah tersebut dengan menghadirkan pendekatan yang tidak hanya teknis dan akurat, tetapi juga kontekstual dan relevan secara sosial.

2. METODE PENELITIAN

Adapun alur dalam penelitian ini mencakup 6 tahap, yaitu dengan *data collection, pre-processing data, labelling, feature extraction* dengan TF-IDF, classification model menggunakan algoritma Support Vector Machine, dan tahap akhir evaluation model. Alur dari Flowchart penelitian ini dapat dilihat pada Gambar 1 dalam penelitian ini.



Gambar 1. Alur Penelitian

2.1 Data *Collecting*

Data dikumpulkan dari platform media sosial X (sebelumnya Twitter) menggunakan library Python bernama harvest yang dijalankan di Google Colab. Pengambilan dilakukan terhadap tweet berbahasa Indonesia dengan kata kunci "Korupsi Timah" selama periode Maret hingga Mei 2024. Jumlah total data yang berhasil dikumpulkan dan digunakan dalam penelitian ini adalah sebanyak 535 tweet X[9].

Tabel 1. Dataset Media Sosial X

No	Text
0	Kalo pak prabowo itu bener ngomongnya coba sel...
1	Helena Lim yang menjadi bagian maling timah 27...
2	Hrsnya para tersangka 271 T korupsi PT Timah k...
3	Operasi penyitaan lalu digelar. Menurut penghi...
4	@misskadrun_ Antum juga salut sama yg pada kor...

2.2 *Pre-Processing* Data

Bertujuan untuk membersihkan data teks agar siap digunakan dalam proses pelatihan model. Tahapan yang dilakukan meliputi *case folding*[10] (mengubah seluruh huruf menjadi kecil), *cleansing* (menghapus tanda baca, angka, URL, simbol, dan data kosong), *tokenization* (memecah kalimat menjadi token kata), *stopword removal* (menghilangkan kata-kata umum yang tidak relevan), serta *stemming* (mengembalikan kata ke bentuk dasarnya dengan library Sastrawi).

2.3 *Labelling*

Dilakukan secara otomatis menggunakan *library TextBlob*[8]. Tweet diklasifikasikan ke dalam dua kategori sentimen: positif dan negatif. Dari total 535 tweet, sebanyak 516 dikategorikan sebagai negatif dan 19 sebagai positif, menunjukkan dominasi opini publik yang cenderung negatif terhadap kasus yang diteliti.

2.4 *Feature Extraction*

Menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah data teks menjadi representasi numerik[11]. Metode ini digunakan untuk memberikan bobot yang mencerminkan pentingnya suatu kata dalam sebuah dokumen terhadap keseluruhan korupsi.

2.5 Penerapan *Classification* Support Vector Machine

Data yang telah direpresentasikan dalam bentuk vektor TF-IDF dibagi menjadi dua bagian menggunakan metode dengan rasio 80% data latih (428 tweet) dan 20% data uji (107 tweet). Pembagian ini dilakukan secara acak dengan random_state=42 untuk menjaga reprodusibilitas. Setelah data vektor TF-IDF dibagi menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji[12]. Algoritma Support Vector Machine dipilih karena kemampuannya dalam menangani klasifikasi teks berdimensi tinggi serta memaksimalkan margin antar kelas[13]. Model diimplementasikan menggunakan bahasa pemrograman Python dalam lingkungan Google Colab.

2.6 *Evaluation Model*

Menggunakan metrik akurasi, *precision, recall, F1-score*, serta *confusion matrix*. Hasil evaluasi menunjukkan bahwa model SVM memberikan kinerja yang sangat baik dengan akurasi mencapai 98%, serta nilai precision dan recall yang tinggi terutama dalam mengidentifikasi sentimen negatif[14]. Hasil ini membuktikan bahwa algoritma SVM efektif dalam klasifikasi sentimen berbasis data media sosial.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset

Penelitian ini dilakukan dengan memanfaatkan data opini publik yang berkaitan langsung dengan isu korupsi dalam sektor pertambangan timah, yang merupakan salah satu permasalahan yang sedang menjadi perhatian masyarakat luas. Data yang digunakan dalam penelitian ini diperoleh dari berbagai sumber digital, termasuk namun tidak terbatas pada media sosial, portal berita daring, serta platform diskusi publik yang secara aktif digunakan masyarakat untuk menyuarakan pandangan, tanggapan, dan pendapat mereka terhadap permasalahan yang berkembang di tengah masyarakat. Data opini publik ini sangat relevan mengingat bahwa media sosial dan ruang digital saat ini telah menjadi wadah utama dalam penyebaran informasi serta pembentukan persepsi publik, termasuk terhadap isu-isu krusial seperti tindak pidana korupsi di sektor strategis negara, dalam hal ini adalah pertambangan timah.

```
# ----------------------------------------
# 2. Load Dataset
# ----------------------------------------
print("=== Membaca Dataset ===")
df = pd.read_csv('/content/Data_Korupsi_Timah_271T.csv')
df['text'] = df['full_text'].astype(str)
print(f"Jumlah data: {len(df)}")
print("Contoh 5 data awal:\n", df['text'].head())
```

```
=== Membaca Dataset ===
Jumlah data: 535
Contoh 5 data awal:
 0    Kalo pak prabowo itu bener ngomongnya coba sel...
1    Helena Lim yang menjadi bagian maling timah 27...
2    Hrsnya para tersangka 271 T korupsi PT Timah k...
3    Operasi penyitaan lalu digelar. Menurut penghi...
4    @misskadrun_ Antum juga salut sama yg pada kor...
Name: text, dtype: object
```

Gambar 2. Data Penelitian

Dataset yang diperoleh kemudian dianalisis menggunakan pendekatan analisis sentimen, yaitu metode yang bertujuan untuk mengidentifikasi dan mengklasifikasikan sikap atau opini seseorang terhadap suatu topik berdasarkan teks yang diungkapkan. Dalam konteks penelitian ini, analisis sentimen digunakan[15].

Untuk mengetahui kecenderungan atau polaritas opini masyarakat, apakah opini tersebut bersifat positif—yang menunjukkan dukungan, kepercayaan, atau optimisme—atau bersifat negatif, yang mencerminkan rasa kecewa, marah, sinis, ataupun kritik terhadap isu korupsi timah yang sedang hangat diperbincangkan. Dataset akhir yang digunakan dalam penelitian ini berjumlah 535 tweet, yang seluruhnya telah melalui tahapan pra-pemrosesan dan pelabelan sentimen secara otomatis menggunakan pendekatan berbasis leksikon. Dari total ini, 428 tweet digunakan sebagai data latih dan 107 tweet sebagai data uji. Setiap data teks tersebut kemudian diberi label sentimen secara otomatis dengan menggunakan

pendekatan berbasis leksikon (*lexicon-based*), yaitu teknik yang memanfaatkan kamus kata atau daftar leksikal yang telah dipetakan berdasarkan polaritas emosional nya untuk menentukan jenis sentimen dari suatu kalimat atau teks.

Berdasarkan hasil dari proses labelling menggunakan pendekatan tersebut, diketahui bahwa dari total 107 data yang dianalisis, sebanyak 102 data diklasifikasikan ke dalam kategori sentimen negatif, sementara hanya terdapat 5 data yang termasuk dalam kategori sentimen positif. Ketimpangan distribusi ini mengindikasikan bahwa dominasi opini publik terhadap isu korupsi timah sangat kuat berada pada sisi negatif. Hal ini mencerminkan tingginya tingkat kekecewaan, ketidakpuasan, serta kecaman dari masyarakat terhadap praktik korupsi yang terjadi dalam sektor pertambangan yang seharusnya menjadi sumber daya strategis dan memberikan manfaat besar bagi kesejahteraan negara dan rakyat. Opini negatif ini dapat berisi ekspresi marah terhadap pelaku korupsi, sindiran kepada pemerintah atau aparat penegak hukum, serta bentuk kekhawatiran terhadap masa depan pengelolaan sumber daya alam.

```
# 3. Labeling Sentimen (Lexicon-Based)
# ----------------------------------------
print("\n=== Proses Labeling Sentimen (Lexicon-Based) ===")
def get_sentiment(text):
    blob = TextBlob(text)
    return 'positif' if blob.sentiment.polarity > 0 else 'negatif'

df['label'] = df['text'].apply(get_sentiment)
print("Distribusi label:\n", df['label'].value_counts())
print("Contoh data setelah pelabelan:\n", df[['text', 'label']].head())
```

```
=== Proses Labeling Sentimen (Lexicon-Based) ===
Distribusi label:
 label
negatif    516
positif     19
Name: count, dtype: int64
Contoh data setelah pelabelan:
                                                text    label
0  Kalo pak prabowo itu bener ngomongnya coba sel...  negatif
1  Helena Lim yang menjadi bagian maling timah 27...  negatif
2  Hrsnya para tersangka 271 T korupsi PT Timah k...  negatif
3  Operasi penyitaan lalu digelar. Menurut penghi...  negatif
4  @misskadrun_ Antum juga salut sama yg pada kor...  negatif
```

Gambar 3. *Labelling* data

Sementara itu, opini positif yang sangat sedikit jumlahnya mungkin mencerminkan harapan atau dukungan terhadap langkah pemberantasan korupsi atau terhadap tokoh atau institusi yang dianggap sedang berupaya menyelesaikan permasalahan tersebut. Ketidakseimbangan jumlah opini ini juga memberikan tantangan tersendiri dalam proses klasifikasi, karena model analisis sentimen perlu mampu mengatasi kondisi distribusi kelas yang tidak seimbang atau disebut juga dengan istilah *class imbalance*, yang dapat memengaruhi performa klasifikasi dalam mengenali sentimen minoritas. Oleh karena itu, hasil distribusi ini tidak hanya memberikan informasi tentang persepsi publik secara umum, tetapi juga memberikan dasar yang kuat bagi tahapan selanjutnya dalam proses pemodelan klasifikasi dan evaluasi kinerja algoritma yang digunakan dalam penelitian ini.

3.2 Proses Pra-pemrosesan Data

Pra-pemrosesan data merupakan tahap penting dalam analisis teks, terutama saat menggunakan algoritma machine learning untuk klasifikasi. Kualitas model sangat bergantung

pada kualitas data yang digunakan. Data mentah dari media sosial umumnya tidak rapi, sering berisi huruf kapital acak, tanda baca tidak konsisten, emoticon, dan singkatan informal. Oleh karena itu, data perlu dibersihkan dan dinormalisasi agar siap untuk dianalisis.

Tahapan pertama adalah cleaning, yaitu menghapus karakter non-alfabet seperti simbol, angka, dan tautan yang tidak relevan dengan analisis sentimen. Selanjutnya dilakukan lowercasing untuk menyamakan semua huruf menjadi kecil, sehingga kata seperti “Korupsi” dan “korupsi” dianggap sama.

```
# ----------------------------------------
# 4. Pra-pemrosesan Data Teks
# ----------------------------------------
print("\n=== Pra-pemrosesan Teks ===")
def clean_text(text):
    text = text.lower()
    text = re.sub(r"http\S+", "", text)  # hapus URL
    text = re.sub(r"@\w+", "", text)     # hapus mention
    text = re.sub(r"#\w+", "", text)     # hapus hashtag
    text = re.sub(r"[^a-zA-Z\s]", "", text)  # hapus simbol/angka
    text = re.sub(r"\s+", " ", text)     # hapus spasi ganda
    return text.strip()

df['clean_text'] = df['text'].apply(clean_text)
print("Contoh hasil pra-pemrosesan:\n", df[['text', 'clean_text']].head())
```

```
=== Pra-pemrosesan Teks ===
Contoh hasil pra-pemrosesan:
                                                text  \
0  Kalo pak prabowo itu bener ngomongnya coba sel...
1  Helena Lim yang menjadi bagian maling timah 27...
2  Hrsnya para tersangka 271 T korupsi PT Timah k...
3  Operasi penyitaan lalu digelar. Menurut penghi...
4  @misskadrun_ Antum juga salut sama yg pada kor...

                                          clean_text
0  kalo pak prabowo itu bener ngomongnya coba sel...
1  helena lim yang menjadi bagian maling timah t ...
2  hrsnya para tersangka t korupsi pt timah kmr j...
3  operasi penyitaan lalu digelar menurut penghit...
4  antum juga salut sama yg pada korupsi timah ke...
```

Gambar 4. Pembersihan data Penelitian

Proses memecah kalimat atau paragraf menjadi potongan-potongan kata atau unit terkecil yang disebut token. Tokenisasi sangat penting karena memungkinkan sistem untuk mengenali setiap kata sebagai entitas yang dapat dianalisis secara individual. Setelah itu, dilakukan penghapusan stopword (*stopword removal*), yakni proses mengeliminasi kata-kata yang secara linguistik sering muncul dalam bahasa Indonesia namun tidak memberikan kontribusi signifikan terhadap penentuan makna atau sentimen, seperti kata “yang”, “dan”, “di”, dan sebagainya. Kata-kata ini umumnya tidak membawa bobot informasi yang penting dalam konteks analisis sentimen.

Tahapan terakhir dalam proses pra-pemrosesan adalah stemming, yaitu proses untuk mengubah setiap kata ke bentuk dasarnya. Misalnya, kata “menyuarakan”, “suara”, dan “bersuara” akan dikembalikan ke bentuk dasar “suara”. Proses stemming dilakukan menggunakan stemmer yang sesuai dengan bahasa Indonesia untuk memastikan bahwa model pembelajaran mesin dapat mengenali kata-kata dengan akar yang sama sebagai satu kesatuan makna, sehingga mengurangi redundansi dan meningkatkan akurasi pemodelan. Dengan melalui seluruh tahapan pra-pemrosesan ini, data teks yang semula tidak terstruktur dan berantakan diubah menjadi bentuk yang lebih seragam dan bermakna, sehingga dapat meningkatkan performa algoritma klasifikasi yang akan digunakan dalam tahap selanjutnya, yaitu implementasi Support Vector Machine (SVM) dalam analisis sentimen.

3.3 Implementasi Support Vector Machine (SVM)

Setelah melalui tahapan pra-pemrosesan, data siap digunakan dalam proses klasifikasi. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) yang dikenal efektif

dalam klasifikasi teks berdimensi tinggi. SVM bekerja dengan mencari hyperplane optimal yang memisahkan dua kelas dengan margin maksimum, sehingga cocok untuk membedakan opini publik menjadi sentimen positif atau negatif.

Tabel 2. Hasil TF-IDF

Term	TF-IDF
rakyat	0.3158
aja	0.3109
adiknya	0.2205
hartanya	0.2205
hasyim	0.2205
ngomongnya	0.2205
miskin	0.207
pak	0.207
smelter	0.207
berani	0.1973
demi	0.1973
gak	0.1973
malah	0.1973
coba	0.1786
presiden	0.1786
kalo	0.1742
bikin	0.1667
ilegal	0.1635
baru	0.1531
indonesia	0.1531
tambang	0.151
mau	0.147
bagi	0.1374
itu	0.136
ya	0.1335
pada	0.1288
harus	0.1016
korupsi	0.0523

Model ini menggunakan representasi fitur berbasis TF-IDF (*Term Frequency-Inverse Document Frequency*), yang memberikan bobot pada kata-kata berdasarkan frekuensi kemunculan nya dan relevansi nya dalam seluruh dokumen. TF-IDF membantu meningkatkan akurasi dengan menyoroti kata-kata yang memiliki daya pembeda tinggi antar kelas sentimen.

3.4 Hasil Evaluasi Model

Evaluasi model dilakukan untuk menilai kemampuan generalisasi algoritma Support Vector Machine (SVM) yang telah dilatih dengan representasi fitur TF-IDF. Evaluasi mencakup dua komponen utama: *classification report* dan *confusion matrix*, yang menggambarkan performa model dalam mengklasifikasikan data sentimen positif dan negatif.

Tabel 3. Hasil *Report* Klasifikasi

Label	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	0.98	1.00	0.99	102
Positif	1.00	0.60	0.75	5
Accuracy			0.98	107
Macro Avg	0.99	0.80	0.87	107
Weighted Avg	0.98	0.98	0.98	107

Model menunjukkan performa sangat baik dalam mendeteksi sentimen negatif dengan precision 0.98, recall 1.00, dan F1-score 0.99. Untuk sentimen positif, precision mencapai 1.00 namun recall hanya 0.60, yang menunjukkan adanya beberapa data positif yang salah klasifikasi sebagai negatif. Secara keseluruhan, model memperoleh akurasi sebesar 98%, dengan macro average F1-score sebesar 0.87 dan weighted average sebesar 0.98.

Tabel 4. *Confusion Matrix*

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	102	0
Aktual Positif	2	3

Confusion matrix menunjukkan seluruh 102 data negatif diklasifikasi dengan benar, sedangkan dari 5 data positif, 3 diklasifikasi benar dan 2 salah. Ketidakseimbangan jumlah data antar kelas menjadi tantangan utama, sehingga pendekatan seperti resampling atau penyesuaian threshold dapat dipertimbangkan pada studi lanjutan untuk meningkatkan deteksi kelas minoritas.

4. KESIMPULAN

Penelitian ini membahas penerapan algoritma Support Vector Machine (SVM) dalam klasifikasi sentimen masyarakat terhadap kasus korupsi di sektor pertambangan timah, dengan memanfaatkan data dari media sosial X (sebelumnya Twitter). Hasil penelitian menunjukkan bahwa SVM merupakan algoritma yang andal dalam menangani klasifikasi teks berbasis opini publik, terutama saat dikombinasikan dengan metode representasi fitur TF-IDF. Model yang dibangun mampu mencapai tingkat akurasi tinggi, yaitu sebesar 98%, dengan nilai *precision* dan *recall* yang sangat baik pada kelas sentimen negatif. Hal ini menunjukkan bahwa SVM sangat efektif dalam mengidentifikasi opini negatif yang eksplisit secara linguistik. Namun demikian, performa model terhadap kelas sentimen positif masih tergolong rendah, ditunjukkan oleh nilai recall sebesar 0.60. Ini mengindikasikan bahwa sebagian opini positif gagal dikenali dengan benar, kemungkinan besar disebabkan oleh ketidakseimbangan jumlah data antar kelas (*class imbalance*), di mana opini negatif jauh lebih dominan.

Sebagai rekomendasi untuk penelitian selanjutnya, disarankan beberapa langkah nyata yang dapat meningkatkan performa dan keluasan studi:

1. Mengumpulkan data yang lebih besar dan seimbang, khususnya memperbanyak data sentimen positif untuk mengurangi ketimpangan kelas.
2. Menerapkan teknik penyeimbangan data seperti SMOTE (*Synthetic Minority Over-sampling Technique*) atau under sampling pada data mayoritas untuk meningkatkan sensitivitas model terhadap sentimen minoritas.
3. Melakukan perbandingan antar algoritma (misalnya, Naïve Bayes, Random Forest, atau Transformer-based models seperti BERT) untuk mengetahui pendekatan mana yang paling efektif pada konteks opini publik terhadap isu korupsi.
4. Menggunakan pendekatan *deep learning* atau *ensemble learning* yang dapat menangkap konteks linguistik yang lebih kompleks, terutama dalam opini yang implisit atau ambigu.
5. Menambahkan analisis temporal atau spasial, misalnya dengan memetakan dinamika sentimen dari waktu ke waktu atau berdasarkan lokasi geografis (jika metadata tersedia), guna memahami perubahan persepsi publik secara lebih holistik.
6. Dengan langkah-langkah tersebut, penelitian di masa depan diharapkan mampu tidak hanya meningkatkan akurasi klasifikasi, tetapi juga memperkaya wawasan mengenai persepsi publik terhadap isu-isu strategis nasional berbasis analisis data digital.

DAFTAR PUSTAKA

[1] R. R. Harahap and Mhd. Furqan, “Sentiment Analysis towards the 2024 Vice Presidential Candidate Debate Using the Support Vector Machine Algorithm,” *sinkron*, vol. 8, no. 3, pp. 1783–1794, Jul. 2024, doi: 10.33395/sinkron.v8i3.13903.

[2] A. Setiawan and R. R. Suryono, “Edumatic: Jurnal Pendidikan Informatika Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes,” vol. 8, no. 1, pp. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.

[3] A. Handayani and I. Zufria, “Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM,” *Journal of Information System Research (JOSH)*, vol. 5, no. 1, pp. 53–63, Oct. 2023, doi: 10.47065/josh.v5i1.4379.

[4] N. Meilani, Mhd. Furqan, and Suhardi, “Analisis sentimen pengguna aplikasi BSI mobile akibat ransomware menggunakan algoritma support vector machine,” *INFOTECH: Jurnal Informatika & Teknologi*, vol. 5, no. 1, pp. 42–51, Jun. 2024, doi: 10.37373/infotech.v5i1.1102.

[5] M. Furqan and A. ab. N. Fakhri, “Big Data Approach to Sentiment Analysis in Machine Learning-Based Microblogs: Perspectives of Religious Moderation Public Policy in Indonesia,” May 2024.

[6] M. S. Hasibuan and A. Serdano, “Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naive Bayes,” *JRST (Jurnal Riset Sains dan Teknologi)*, vol. 6, no. 2, p. 199, Nov. 2022, doi: 10.30595/jrst.v6i2.15145.

[7] K. L. Ong *et al.*, “Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021,” *The Lancet*, vol. 402, no. 10397, pp. 203–234, Jul. 2023, doi: 10.1016/S0140-6736(23)01301-6.

[8] Y. Laia, S. Sandino Berutu, el Pieter Sumihar, and H. Budiati, “Implementasi Library Textblob dan Metode Support Vector Machine Pada Analisis Sentimen Pelanggan

Terhadap Jasa Transportasi Online," *Technology and Science (BITS)*, vol. 6, no. 1, 2024, doi: 10.47065/bits.v6i1.5090.

[9] M. D. R. Ramadhan and N. Yuliana, "Analisis Pola Perilaku Penggemar JKT48 pada Aplikasi X dalam Berkomunikasi dengan Sesama Penggemar," *Triwikrama: Jurnal Multidisiplin Ilmu Sosial*, vol. 2, no. 4, pp. 31–40, 2023, [Online]. Available: https://doi.org/10.6578/triwikrama.v2i5.1065

[10] A. Brancaccio *et al.*, "Towards the definition of a standard in TMS-EEG data preprocessing," *Neuroimage*, vol. 301, Nov. 2024, doi: 10.1016/j.neuroimage.2024.120874.

[11] R. T. Wahyuni, D. Prastiyanto, and D. E. Supraptono, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi."

[12] N. Huda Ovirianti, M. Zarlis, and H. Mawengkang, "Support Vector Machine Using a Classification Algorithm," *Jurnal dan Penelitian Teknik Informatika*, vol. 6, no. 3, 2022, doi: 10.33395/sinkron.v7i3.

[13] M. Furqan, R. Kurniawan, and K. I. HP, "Evaluasi Performa Support Vector Machine Classifier Terhadap Penyakit Mental," *Jurnal Sistem Informasi Bisnis*, vol. 10, no. 2, pp. 203–210, Dec. 2020, doi: 10.21456/vol10iss2pp203-210.

[14] D. Mualfah, Ramadhoni, R. Gunawan, and D. Mulyadipa Suratno, "Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM," *Jurnal Fasilkom*, vol. 13, no. 01, pp. 72–80, 2023, doi: 10.37859/jf.v13i01.4920.

[15] N. Qomariya, A. Kisnu Darmawan, M. Walid, and Y. Efenie, "Sentiment Analysis on LGBT Issues in Indonesia with Lexicon-Based and Support Vector Machine Algorithms," vol. 19, no. 1, 2023, doi: 10.33480/pilar.v18i1.4183.