



Analisis Sentimen Berita Skandal Korupsi pada Media Sosial X Menggunakan Metode Linear Support Vector Machine Berbasis Android

Villa Dea Ardinata^{1*}, Martiano²

^{1*,2}Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Email: ^{1*}villadeaardinata@gmail.com, ²martiano@umsu.ac.id

Abstrak

Korupsi merupakan permasalahan serius yang berdampak pada keuangan negara, tatanan sosial, dan kepercayaan publik terhadap institusi pemerintahan. Banyaknya opini masyarakat yang tersebar di media sosial X (sebelumnya Twitter) menjadikan analisis sentimen sebagai pendekatan efektif untuk memahami kecenderungan persepsi publik terhadap berita skandal korupsi. Penelitian ini bertujuan menganalisis sentimen publik menggunakan metode Linear Support Vector Machine (SVM) serta mengimplementasikan hasil analisis ke dalam aplikasi berbasis Android. Tahapan penelitian meliputi scraping data menggunakan ntscraper dan API X, preprocessing (cleansing, case folding, tokenization, normalization, stopword removal, dan stemming), pelabelan sentimen secara manual, pembobotan fitur menggunakan TF-IDF, pembagian data latih dan uji, serta evaluasi model menggunakan confusion matrix dan metrik accuracy, precision, recall, dan F1-Score. Penelitian berhasil mengumpulkan 1.011 tweet terkait skandal korupsi pada periode 1 Mei–15 Agustus 2026. Hasil pengujian menunjukkan model Linear SVM mencapai akurasi 89% pada 203 data uji, dengan weighted precision 89%, weighted recall 89%, dan weighted F1-Score 89%. Temuan ini menunjukkan bahwa Linear SVM efektif mengklasifikasikan sentimen pada teks berdimensi tinggi dan layak diterapkan sebagai dasar pengembangan aplikasi analisis sentimen berbasis Android.

Kata Kunci: Analisis Sentimen; Korupsi; Linear Support Vector Machine; Media Sosial X; TF-IDF.

PENDAHULUAN

Korupsi merupakan salah satu permasalahan serius yang masih dihadapi oleh Indonesia. Tindak kejahatan ini tidak hanya merugikan keuangan negara tetapi juga merusak tatanan sosial, mengganggu pembangunan nasional, dan melemahkan kepercayaan publik terhadap institusi pemerintahan (Hidayat et al., 2025). Berbagai skandal korupsi yang melibatkan pejabat publik kerap menjadi sorotan luas dan memicu reaksi opini publik secara masif. Kondisi ini menunjukkan bahwa korupsi bukan sekadar persoalan hukum, melainkan juga persoalan sosial yang memengaruhi persepsi dan sikap masyarakat secara luas (Dewi et al., 2025).

Seiring berkembangnya teknologi informasi, media sosial X (sebelumnya Twitter) telah menjadi ruang publik digital yang signifikan bagi masyarakat untuk mengekspresikan opini, kritik, maupun dukungan terhadap berbagai isu, termasuk berita skandal korupsi. Setiap komentar atau unggahan pengguna di media sosial X mengandung nilai sentimen yang mencerminkan pandangan publik (Muflih et al., 2021). Oleh karena itu, analisis terhadap data teks pada media sosial X menjadi penting untuk mengetahui kecenderungan opini publik—apakah lebih dominan bersentimen positif atau negatif terhadap berita skandal korupsi.

Analisis sentimen adalah proses mengidentifikasi dan mengekstrak opini atau perasaan dari sebuah teks untuk mengenali persepsi publik terhadap suatu topik atau isu tertentu (Ma'rufudin & Yudhistira, 2025). Dalam prosesnya, analisis sentimen mengklasifikasikan teks ke dalam kategori positif atau negatif berdasarkan opini yang terkandung di dalamnya. Pendekatan ini memungkinkan pemrosesan ribuan data tweet secara otomatis untuk memahami kecenderungan opini publik.

Dalam pemilihan metode klasifikasi, Support Vector Machine (SVM) dikenal sebagai salah satu metode yang efektif dalam bidang machine learning karena kemampuannya menghasilkan akurasi tinggi pada berbagai jenis data (Alfatah & Cahyono, 2025). Namun, SVM non-linear yang menggunakan fungsi kernel seperti Radial Basis Function (RBF), sigmoid, dan polynomial memiliki kelemahan berupa

kompleksitas komputasi yang tinggi, waktu pelatihan yang lebih lama, dan risiko overfitting pada data teks berdimensi tinggi seperti postingan media sosial. Beberapa penelitian terdahulu yang menggunakan SVM non-linear pada analisis sentimen masih menghasilkan akurasi yang relatif rendah; misalnya, Alfataah & Cahyono (2025) memperoleh akurasi 79,73% dalam menganalisis sentimen pengguna Twitter terhadap provider XL Axiata, yang disebabkan oleh ketidakcocokan antara dataset dan fungsi kernel yang digunakan.

Berbeda dengan varian non-linearnya, Linear SVM bekerja dengan membangun hyperplane linear tanpa menggunakan fungsi kernel, sehingga proses komputasi lebih cepat, model lebih sederhana, serta performa lebih stabil dengan risiko overfitting yang lebih rendah—terutama pada dataset teks berdimensi tinggi yang umum dijumpai di media sosial (Dasmase et al., 2022). Karakteristik ini menjadikan Linear SVM lebih sesuai untuk klasifikasi sentimen teks media sosial yang direpresentasikan dalam bentuk vektor TF-IDF berdimensi tinggi.

Kajian terhadap penelitian terdahulu menunjukkan bahwa penggunaan kernel linear pada SVM secara konsisten menghasilkan performa yang baik. Styawati et al. (2021) memperoleh akurasi 98% menggunakan kernel linear SVM pada analisis sentimen program Kartu Prakerja di Twitter; Ernawati & Wati (2024) memperoleh akurasi 92,23% dalam analisis sentimen ulasan aplikasi ChatGPT; serta Nugroho & Norhikmah (2024) memperoleh akurasi 89% pada analisis sentimen terkait Covid-19. Konsistensi performa kernel linear pada berbagai domain analisis sentimen memperkuat kelayakannya untuk diterapkan pada konteks berita skandal korupsi.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk: (1) menganalisis sentimen publik terhadap berita skandal korupsi pada media sosial X menggunakan metode Linear SVM; (2) menerapkan metode Linear SVM untuk klasifikasi sentimen berita skandal korupsi; (3) mengukur performa model berdasarkan nilai accuracy, precision, recall, dan F1-Score; serta (4) mengimplementasikan hasil analisis ke dalam aplikasi berbasis Android. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem pemantauan opini publik yang efisien dan dapat diakses secara mobile.

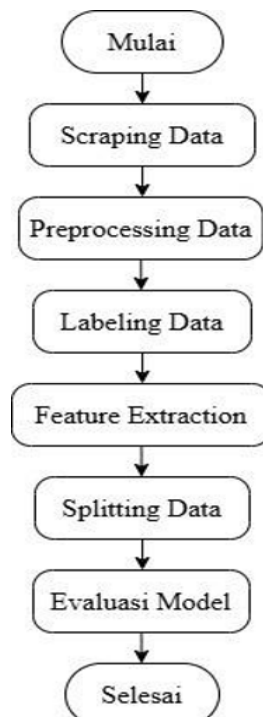
METODE

Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian eksperimental komputasional. Data teks dikumpulkan dari media sosial X kemudian dianalisis secara statistik melalui serangkaian tahapan pemrosesan bahasa alami dan klasifikasi machine learning.

Alur Penelitian

Penelitian dilaksanakan melalui tahapan sistematis sebagai berikut: (1) scraping data, (2) preprocessing data, (3) labeling data, (4) feature extraction menggunakan TF-IDF, (5) splitting data, (6) pelatihan dan evaluasi model Linear SVM, serta (7) implementasi aplikasi Android.



Gambar 1. Alur Penelitian

Pengumpulan Data

Data dikumpulkan dari media sosial X menggunakan teknik scraping dengan library Python ntscraper dan API X. Kata kunci yang digunakan meliputi: "korupsi desa merah putih", "korupsi MBG Glory Harimas Sihombing", "korupsi Chromebook Nadiem Makarim", "korupsi timah", "korupsi bansos", "harvey moeis", dan variasi kata kunci terkait skandal korupsi lainnya. Pengumpulan data dilaksanakan pada periode 1 Mei–15 Agustus 2026 dan menghasilkan 1.011 tweet berbahasa Indonesia. Data yang terkumpul disimpan dalam format CSV untuk pemrosesan selanjutnya.

Preprocessing Data

Preprocessing bertujuan meningkatkan kualitas data teks sebelum digunakan dalam pemodelan. Tahapan preprocessing yang diterapkan adalah sebagai berikut:

1. **Cleansing:** Pembersihan data dari elemen tidak relevan, meliputi karakter HTML, tautan URL, tagar (#), mention pengguna (@username), angka, emoticon, dan karakter khusus lainnya yang berpotensi menimbulkan noise dalam analisis.
2. **Case Folding:** Penyeragaman seluruh huruf dalam teks menjadi huruf kecil (lowercase) guna menghindari inkonsistensi saat proses tokenisasi.
3. **Tokenization:** Pemecahan kalimat menjadi unit-unit kata (token) berdasarkan spasi dan tanda baca, sehingga memudahkan pemrosesan pada tahap selanjutnya.
4. **Normalization:** Penyeragaman kata tidak baku, singkatan, dan ejaan tidak formal yang lazim digunakan di media sosial menjadi bentuk kata baku sesuai Kamus Besar Bahasa Indonesia (KBBI).
5. **Stopword Removal:** Penghapusan kata-kata yang dianggap tidak bermakna dalam analisis sentimen, seperti kata keterangan, kata sambung, dan kata penghubung.
6. **Stemming:** Pengubahan kata berimbuhan (awalan dan akhiran) menjadi kata dasar menggunakan algoritma Sastrawi, yang dirancang khusus untuk bahasa Indonesia.

Labeling Data

Pelabelan sentimen dilakukan secara manual pada seluruh data tweet yang telah melalui tahap preprocessing. Setiap tweet diberi label: nilai **0** untuk sentimen negatif dan nilai **1** untuk sentimen positif. Proses ini dilakukan dengan mempertimbangkan konteks kalimat secara keseluruhan, bukan sekadar kata individual.

Feature Extraction: Pembobotan TF-IDF

Representasi numerik data teks dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen serta kelangkaannya di seluruh koleksi dokumen, sehingga kata-kata yang lebih informatif mendapat bobot lebih tinggi (Septiani & Isabela, 2022). Rumus perhitungan TF-IDF adalah sebagai berikut:

1. Term Frequency (TF)

$$TF(t, d) = \frac{f_{t,d}}{\max_{t' \in d} f_{t',d}}$$

di mana $f_{t,d}$ adalah jumlah kemunculan kata t pada dokumen d , dan $\max_{t' \in d} f_{t',d}$ adalah frekuensi kemunculan tertinggi dari seluruh kata dalam dokumen yang sama.

2. Inverse Document Frequency (IDF)

$$IDF(t, D) = \log_2 \left(\frac{N}{df_t} \right)$$

di mana N merupakan total dari jumlah dokumen dan df_t adalah jumlah dokumen yang mengandung kata t .

3. Bobot TF-IDF

$$W(t, d) = TF(t, d) \times IDF(t, D) \quad \dots (1)$$

Splitting Data

Dataset berlabel dibagi menjadi data latih (*training set*) dan data uji (*testing set*) dengan rasio **80:20**. Dari total 1.011 data, sebanyak 808 data digunakan untuk melatih model dan 203 data digunakan untuk menguji performa model.

Pelatihan Model Linear SVM

Linear SVM adalah algoritma klasifikasi yang bekerja dengan mencari hyperplane linear optimal untuk memisahkan data ke dalam kelas yang berbeda (Dasmase et al., 2022). Berbeda dengan SVM berbasis kernel, Linear SVM tidak menggunakan fungsi transformasi ruang fitur, sehingga komputasinya lebih efisien khususnya pada data bertipe vektor TF-IDF berdimensi tinggi seperti data teks media sosial. Model dilatih menggunakan data latih yang telah direpresentasikan sebagai vektor TF-IDF.

Evaluasi Model

Performa model dievaluasi menggunakan confusion matrix serta empat metrik evaluasi standar:

Tabel 1. Confusion Matrix

	Prediksi Positif	Prediksi Negatif
Aktual Positif	True Positive (TP)	False Negative (FN)
Aktual Negatif	False Positive (FP)	True Negative (TN)

Keterangan:

TP = data positif yang diprediksi positif;

TN = data negatif yang diprediksi negatif;

FP = data negatif yang diprediksi positif;

FN = data positif yang diprediksi negatif.

Metrik evaluasi dihitung dengan rumus berikut (Aisah et al., 2023):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \dots (2)$$

$$Precision = \frac{TP}{TP + FP} \quad \dots (3)$$

$$Recall = \frac{TP}{TP + FN} \quad \dots (4)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \dots (5)$$

Implementasi Aplikasi Android

Hasil model Linear SVM diintegrasikan ke dalam aplikasi berbasis Android yang dikembangkan menggunakan Android Studio dengan bahasa pemrograman Java. Proses machine learning dilakukan menggunakan Python, sedangkan Firebase digunakan sebagai platform penyimpanan dan sinkronisasi data secara real-time. Arsitektur sistem dirancang menggunakan Unified Modeling Language (UML) yang mencakup use case diagram dan activity diagram.

HASIL DAN PEMBAHASAN

Hasil Scraping Data

Proses scraping menggunakan library Python ntscraper dan API X berhasil mengumpulkan **1.011 tweet** berbahasa Indonesia terkait berita skandal korupsi pada periode 1 Mei–15 Agustus 2026. Kata kunci yang digunakan mencakup berbagai isu korupsi aktual, antara lain "korupsi desa merah putih", "korupsi MBG Glory Harimas Sihombing", dan "korupsi Chromebook Nadiem Makarim". Data scraping disimpan dalam format CSV yang berisi konten tweet beserta metadata relevan.

```

from ntscraper import Nitter
import pandas as pd
import os

def crawl_x_data(keywords, limit=50):
    scraper = Nitter()
    all_tweets = []

```

	date	username	content	stemmed_text	predicted_sentiment
0	Thu, 13 Aug 2026 23:02:00 GMT	kompasscom	Berita terbaru mengenai korupsi desa merah put...	berita baru kena korupsi desa merah putih dari...	
1	Sun, 02 Mar 2026 08:00:00 GMT	kompasscom	Berita terbaru mengenai korupsi digitalisasi s...	berita baru kena korupsi digitalisasi spbu per...	
2	Fri, 14 Aug 2026 10:43:00 GMT	kompasscom	Berita terbaru mengenai bupati kuantan singing...	berita baru kena bupati antan singingi suhard...	
3	Tue, 30 Jun 2026 07:00:00 GMT	kompasscom	Berita terbaru mengenai korupsi chromebook nad...	berita baru kena korupsi chromebook nadiem mak...	
4	Wed, 12 Aug 2026 10:15:00 GMT	kompasscom	Berita terbaru mengenai kasus korupsi DJKA sud...	berita baru kena kasus korupsi djka sudewo bup...	

Gambar 2. Implementasi dan Hasil Scraping Data

Hasil Preprocessing Data

Preprocessing dilakukan secara bertahap terhadap 1.011 tweet yang terkumpul:

Case Folding dan Cleansing

Seluruh karakter huruf diubah menjadi huruf kecil, kemudian dilanjutkan dengan pembersihan data dari URL, tagar (#), mention (@), angka, emoticon, dan karakter khusus lainnya yang tidak relevan bagi analisis sentimen.

```

# Tahap 1: Case Folding & Cleansing
case_folded = raw_text.lower()
case_foldeds.append(case_folded)

cleansing = re.sub(r'http[s+|www\S+|https\S+', '', case_folded, flags=re.MULTILINE)
cleansing = re.sub(r'@w+|#w+', '', cleansing)
cleansing = re.sub(r'[a-zs]', '', cleansing)
cleansing = re.sub(r'[s+]', ' ', cleansing).strip()
cleansings.append(cleansing)

```

	text	case_folded	cleansing
0	Berita terbaru mengenai korupsi desa merah put...	berita terbaru mengenai korupsi desa merah put...	berita terbaru mengenai korupsi desa merah put...
1	Berita terbaru mengenai korupsi digitalisasi s...	berita terbaru mengenai korupsi digitalisasi s...	berita terbaru mengenai korupsi digitalisasi s...
2	Berita terbaru mengenai bupati kuantan singing...	berita terbaru mengenai bupati kuantan singing...	berita terbaru mengenai bupati kuantan singing...

Gambar 3. Implementasi dan Hasil Case Folding & Cleansing

Tokenization dan Normalization

Teks yang telah bersih dipecah menjadi token-token kata, kemudian kata-kata tidak baku dan singkatan khas media sosial dinormalisasi ke bentuk baku KBBI.

```
# Tahap 2: Tokenization & Normalization
tokens = cleaned.split()
tokens_list.append(str(tokens))

norm_dict = {'yg': 'yang', 'dgn': 'dengan', 'kl': 'kalau', 'ak': 'aku', 'ntu': 'itu', 'ga': 'tidak', 'gak':
normalized = [norm_dict.get(t, t) for t in tokens]
normalized_list.append(str(normalized))
```

Tabel 2: Tokens -> Normalized

	tokens	normalized
0	['berita', 'terbaru', 'mengenai', 'korupsi', '...]	['berita', 'terbaru', 'mengenai', 'korupsi', '...]
1	['berita', 'terbaru', 'mengenai', 'korupsi', '...]	['berita', 'terbaru', 'mengenai', 'korupsi', '...]
2	['berita', 'terbaru', 'mengenai', 'bupati', 'k...]	['berita', 'terbaru', 'mengenai', 'bupati', 'k...]

Gambar 4. Implementasi dan Hasil Tokenization dan Normalization

Stopword Removal

Kata-kata yang tidak berkontribusi pada analisis sentimen, seperti kata sambung dan kata keterangan, dihapus dari token.

```
# Tahap 3: Stopwords Removal
stopwords_removed = [t for t in normalized if t not in stopwords]
stopwords_removed_list.append(str(stopwords_removed))
```

Tabel 3: Normalized -> Stopwords Remove

	normalized	stopwords_removed
0	['berita', 'terbaru', 'mengenai', 'korupsi', '...]	['berita', 'terbaru', 'mengenai', 'korupsi', '...]
1	['berita', 'terbaru', 'mengenai', 'korupsi', '...]	['berita', 'terbaru', 'mengenai', 'korupsi', '...]
2	['berita', 'terbaru', 'mengenai', 'bupati', 'k...]	['berita', 'terbaru', 'mengenai', 'bupati', 'k...]

Gambar 5. Implementasi dan Hasil Stopword Removal

Stemming

Kata-kata berimbuhan diubah menjadi kata dasar menggunakan algoritma Sastrawi, sehingga variasi morfologi kata dapat diseragamkan.

```
# Tahap 4: Stemming
stemmed_text = stemmer.stem(" ".join(stopwords_removed))
stemmed_list.append(stemmed_text)
```

Tabel 4: Stopwords Removed -> Stemmed

	stopwords_removed	stemmed
0	['berita', 'terbaru', 'mengenai', 'korupsi', '...]	berita baru kena korupsi desa merah putih komp...
1	['berita', 'terbaru', 'mengenai', 'korupsi', '...]	berita baru kena korupsi digitalisasi spbu per...
2	['berita', 'terbaru', 'mengenai', 'bupati', 'k...]	berita baru kena bupati antan singingi suhardi...

Gambar 6. Implementasi dan Hasil Stemming

Hasil Labeling Data

Pelabelan dilakukan secara manual terhadap seluruh 1.011 tweet. Dari proses ini diperoleh distribusi sentimen sebagai berikut: 673 tweet bersentimen negatif (66,6%) dan 338 tweet bersentimen positif (33,4%). Distribusi ini menunjukkan bahwa opini publik di media sosial X cenderung lebih negatif terhadap berita skandal korupsi, yang relevan dengan kondisi sosial-politik terkini di Indonesia.

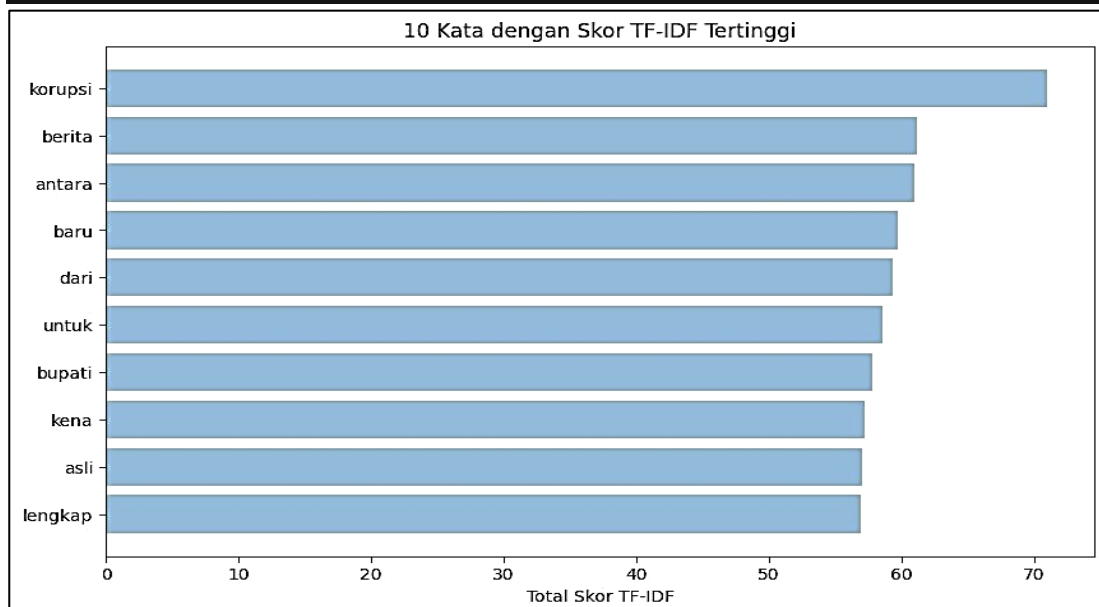
```
=== RINGKASAN DISTRIBUSI SENTIMEN ===  
predicted_sentiment  
0      673  
1      338  
Name: count, dtype: int64  
  
Total data: 1011
```

Gambar 7. Distribusi Sentimen Hasil Labeling Data

Hasil Pembobotan TF-IDF

Representasi numerik data dilakukan menggunakan metode TF-IDF. Proses ini menghasilkan matriks fitur yang merepresentasikan bobot setiap kata dalam masing-masing dokumen tweet secara proporsional terhadap koleksi dokumen keseluruhan.

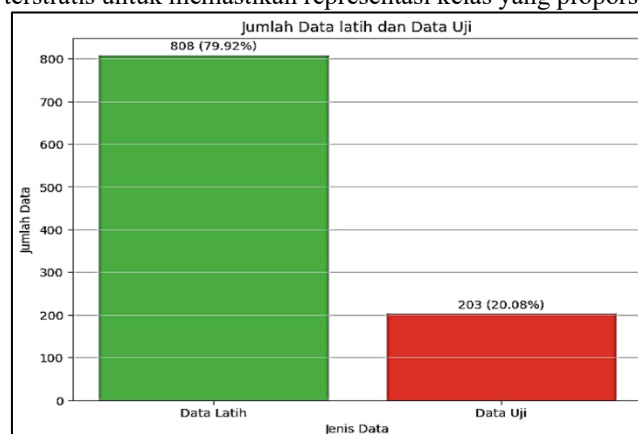
```
# TF-IDF Bar Chart  
print("\n=== MENGHASILKAN GRAFIK TF-IDF ===")  
# Hitung TF-IDF dari seluruh data  
tfidf_vec = TfidfVectorizer(max_features=1000)  
X_tfidf = tfidf_vec.fit_transform(X)
```



Gambar 8. Implementasi TF-IDF dan Skor TF-IDF

Hasil Splitting Data

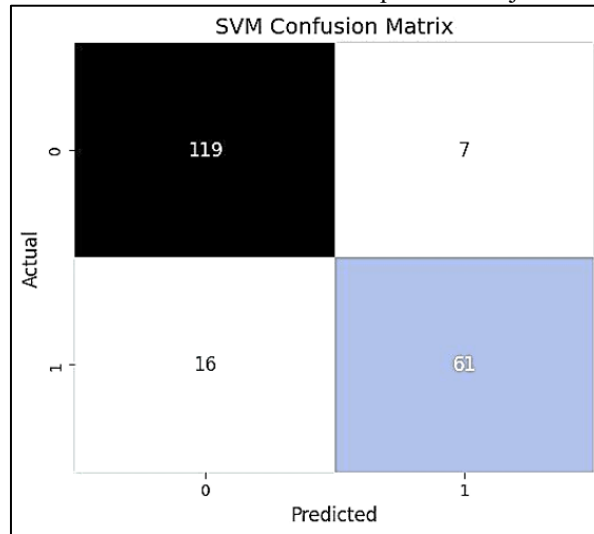
Dataset dibagi dengan rasio 80:20 menghasilkan 808 data latih dan 203 data uji. Pembagian dilakukan secara acak terstratis untuk memastikan representasi kelas yang proporsional pada kedua subset.



Gambar 9. Distribusi Data Latih dan Data Uji

Evaluasi Model Confusion Matrix

Hasil confusion matrix model Linear SVM terhadap 203 data uji adalah sebagai berikut:



Gambar 10. Confusion Matrix Model Linear SVM

Tabel 2. Hasil Confusion Matrix Model Linear SVM

	Prediksi Positif	Prediksi Negatif	Total Aktual
Aktual Positif	61 (TP)	16 (FN)	77
Aktual Negatif	7 (FP)	119 (TN)	126
Total	68	135	203

Berdasarkan hasil confusion matrix, dari 203 data uji: sebanyak 61 data kelas positif diprediksi dengan benar sebagai positif (TP), 119 data kelas negatif diprediksi dengan benar sebagai negatif (TN), 7 data kelas negatif salah diprediksi sebagai positif (FP), dan 16 data kelas positif salah diprediksi sebagai negatif (FN).

Perhitungan Metrik Evaluasi

Berdasarkan nilai confusion matrix di atas, dihitung metrik evaluasi secara manual:

1. Accuracy:

$$Accuracy = \frac{61 + 119}{61 + 119 + 7 + 16} = \frac{180}{203} \approx 88,67\% \approx 89\%$$

2. Precision per kelas:

$$\text{Kelas positif: Precision} = \frac{61}{61+7} = \frac{61}{68} \approx 89,70\% \approx 90\%$$

$$\text{Kelas negatif: Precision} = \frac{119}{119+16} = \frac{119}{135} \approx 88,15\% \approx 88\%$$

3. Recall per kelas:

$$\text{Kelas positif: Recall} = \frac{61}{61+16} = \frac{61}{77} \approx 79,22\% \approx 79\%$$

$$\text{Kelas negatif: Recall} = \frac{119}{119+7} = \frac{119}{126} \approx 94,44\% \approx 94\%$$

4. F1-Score per kelas:

$$\text{Kelas positif: F1} = \frac{2 \times 0,90 \times 0,79}{0,90 + 0,79} \approx 84,13$$

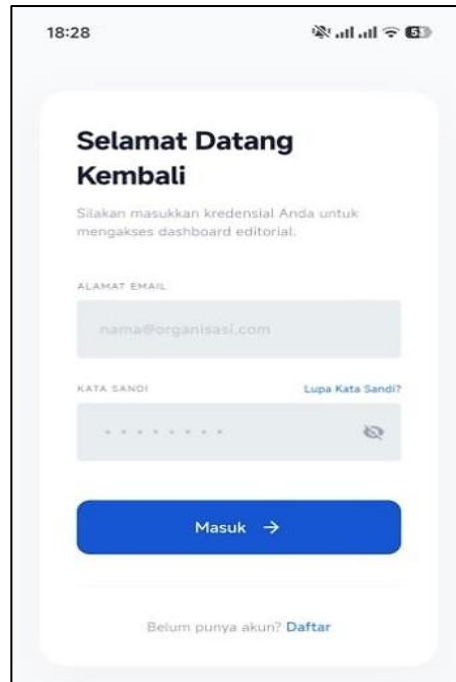
$$\text{Kelas negatif: F1} = \frac{2 \times 0,88 \times 0,94}{0,88 + 0,94} \approx 90,98$$

Classification Report

=== OUTPUT CLASSIFICATION REPORT ===					
Classification Report for SVM:					
	precision	recall	f1-score	support	
0	0.88	0.94	0.91	126	
1	0.90	0.79	0.84	77	
accuracy			0.89	203	
macro avg	0.89	0.87	0.88	203	
weighted avg	0.89	0.89	0.89	203	

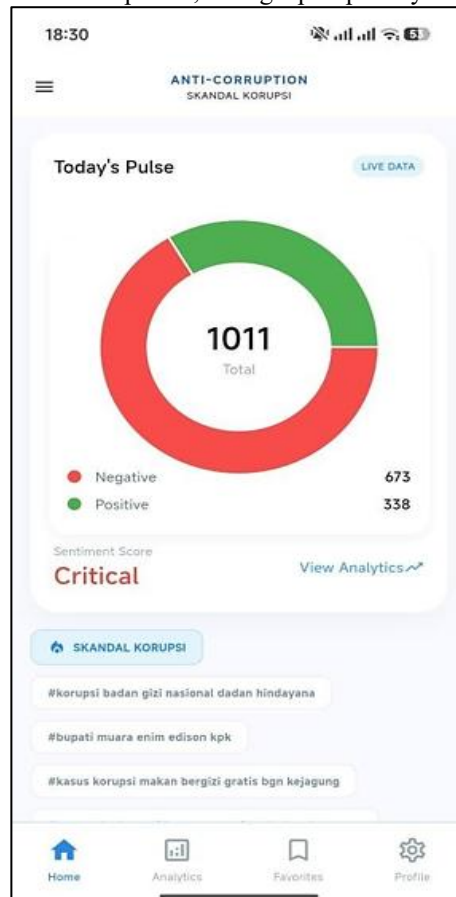
Gambar 11. Hasil Classification Report Model Linear SVM

1. Halaman Login: Halaman autentikasi yang memungkinkan admin mengakses sistem dengan memasukkan email dan password.



Gambar 14. Halaman Login

2. Halaman Home: Menampilkan ringkasan hasil analisis sentimen terhadap 1.011 tweet, dengan rincian 673 sentimen negatif dan 338 sentimen positif, dilengkapi topic keyword dan recent news timeline.



Gambar 15. Halaman Home

3. Halaman Add Keyword: Fitur penambahan topik atau kata kunci baru yang akan dianalisis, dilengkapi kolom nama topik, long-tail queries, rentang periode analisis, dan pengaturan status (Active, Draft, Closed, Archived).

18:29

< Add New Topic

Nama Topik / Keyword *

Masukkan Nama Topik / Keyword

Keywords / Long-tail Queries

Keyword Terkait

Enter keywords or sentences (e.g., 'kerugian negara 271 triliun', 'tindak pidana pencucian uang timah').

Separate multi-sentence keywords with a new line or comma.

Analysis Period

Mulai * Selesai *

yyyy-MM-dd yyyy-MM-dd

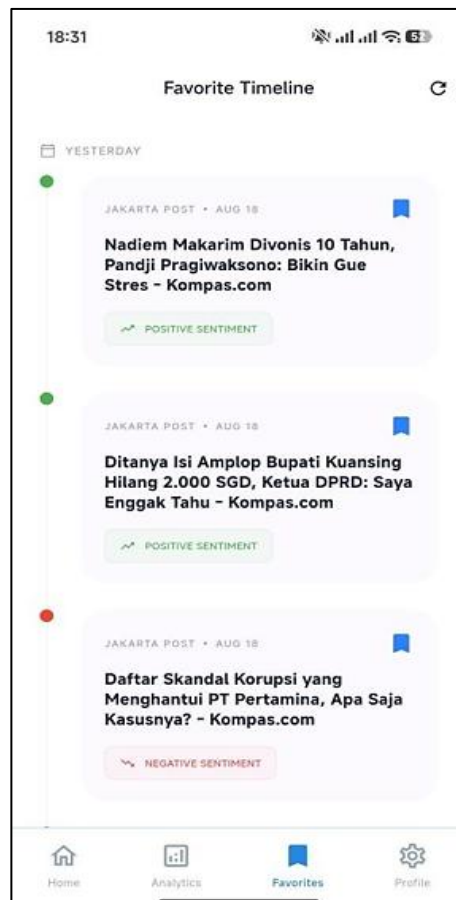
Initial Status

Active Draft Closed Archived

Save Topic

Gambar 16. Halaman Add Keyword

4. Halaman Favorites: Fitur penyimpanan berita skandal korupsi yang ditandai oleh pengguna, ditampilkan dalam format timeline.



Gambar 17. Halaman Favorites

Pembahasan

Hasil penelitian ini mengkonfirmasi efektivitas metode Linear SVM dalam mengklasifikasikan sentimen teks media sosial berbahasa Indonesia. Akurasi sebesar 89% yang diperoleh pada 203 data uji sejalan dengan temuan Nugroho & Norhikmah (2024) yang mencapai akurasi 89% pada analisis sentimen Covid-19, sekaligus melampaui hasil Alfataah & Cahyono (2025) yang menggunakan SVM non-linear dengan kernel RBF dan hanya memperoleh akurasi 79,73%. Hal ini memperkuat argumentasi bahwa Linear SVM lebih sesuai daripada SVM non-linear untuk representasi teks TF-IDF berdimensi tinggi karena data teks yang telah diubah ke vektor TF-IDF pada umumnya sudah bersifat linearly separable di ruang fitur yang berdimensi tinggi, sehingga transformasi kernel tidak memberikan keuntungan signifikan dan justru menambah kompleksitas komputasi.

Weighted average yang menghasilkan precision, recall, dan F1-Score masing-masing sebesar 89% mengindikasikan bahwa model memiliki performa yang seimbang dan tidak bias pada salah satu kelas. Meskipun terdapat perbedaan recall antara kelas negatif (94%) dan kelas positif (79%), perbedaan ini dapat dijelaskan oleh ketidakseimbangan distribusi data (*class imbalance*) di mana data negatif (673 sampel) hampir dua kali lipat data positif (338 sampel). Teknik penyeimbangan data seperti oversampling (SMOTE) atau undersampling pada penelitian lanjutan berpotensi meningkatkan recall kelas positif.

Dari perspektif sosial, dominasi sentimen negatif (66,6%) terhadap berita skandal korupsi mencerminkan ketidakpercayaan dan kekecewaan publik yang signifikan terhadap pelaku korupsi dan sistem penegakan hukum. Temuan ini relevan untuk menjadi masukan bagi pemangku kebijakan dalam memahami persepsi publik dan merancang komunikasi yang lebih efektif dalam upaya pemberantasan korupsi.

Integrasi model ke dalam aplikasi Android memberikan nilai tambah berupa aksesibilitas analisis sentimen secara mobile dan real-time, sehingga memungkinkan pemantauan dinamika opini publik terhadap isu korupsi secara berkelanjutan.

KESIMPULAN

Penelitian ini berhasil menganalisis sentimen publik terhadap berita skandal korupsi di media sosial X menggunakan metode Linear SVM berbasis Android. Proses pengumpulan data melalui *scraping* dengan *library* ntscraper dan API X menghasilkan 1.011 tweet yang kemudian diproses melalui tahapan pra-pemrosesan sehingga siap untuk klasifikasi. Penerapan Linear SVM dengan pembobotan TF-IDF dan pembagian data 80:20 menghasilkan performa yang solid, dengan akurasi 89%, precision dan recall rata-rata tertimbang masing-masing 89%, serta F1-Score 89%. Distribusi sentimen menunjukkan 66,6% negatif dan 33,4% positif, menandakan opini publik cenderung negatif terhadap berita skandal korupsi. Sistem ini berhasil diimplementasikan dalam aplikasi Android dengan fitur *Home*, *Add Keyword*, dan *Favorites* untuk pemantauan sentimen secara real-time.

Saran untuk penelitian selanjutnya, sistem dapat dikembangkan dengan menggunakan dataset yang lebih besar dan periode pengumpulan lebih panjang, menambahkan kategori sentimen netral agar representasi opini lebih komprehensif, serta menerapkan teknik penyeimbangan data seperti SMOTE untuk meningkatkan recall kelas positif. Selain itu, disarankan membandingkan Linear SVM dengan algoritma lain seperti Naïve Bayes, Random Forest, dan Logistic Regression guna memperoleh model dengan performa yang lebih optimal.

DAFTAR PUSTAKA

- Aisah, I. S., Irawan, B., & Suprpti, T. (2023). Algoritma Support Vector Machine (SVM) untuk analisis sentimen ulasan aplikasi Al Qur'an Digital. *Jurnal Teknik Informatika*, 7(6), 3759–3765.
- Alfataah, S., & Cahyono, N. (2025). Sentimen analisis pengguna Twitter terhadap provider XL. *Jurnal Teknik Informatika dan Sistem Informasi*, 9(3), 4502–4506.
- Dasmaseila, R., Latubessy, A., & Susanto, A. (2022). Penerapan metode Support Vector Machine (SVM) untuk mendeteksi penyalahgunaan narkoba. *Journal of Applied Intelligent System*, 1(2), 111–122.
- Dewi, U. K., Malik, S., Yadita, D., Nugraha, R., & Akbar, S. R. (2025). Analisis kasus korupsi bantuan sosial pada pandemi Covid-19 dari perspektif etika dan profesionalisme hukum. *Jurnal Hukum dan Etika Profesi*, 2(6), 59–67.
- Edwin, A. W. S., & Achmad, A. (2018). Pemanfaatan realtime database di platform Firebase pada aplikasi e-tourism Kabupaten Nabire. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 22(1), 20–26. <https://doi.org/10.25042/jpe.052018.04>
- Ernawati, S., & Wati, R. (2024). Evaluasi performa kernel SVM dalam analisis sentimen review aplikasi ChatGPT menggunakan hyperparameter dan VADER lexicon. *Jurnal Ilmiah Sistem Informasi*, 15(April), 40–49.

- Hidayat, S., Azizah, A. N., Angriani, F. S., & Bahtiar, M. A. I. (2025). Tindak pidana korupsi di Indonesia: Bentuk, faktor penyebab, dan dampaknya. *Jurnal Ilmu Hukum*, 3(2), 402–411.
- Ma'rufudin, & Yudhistira, A. (2025). Analisis sentimen petani milenial pada media sosial X menggunakan algoritma Support Vector Machine (SVM). *Jurnal Sistem dan Teknologi Informasi*, 5(3), 845–857.
- Muflih, H. N., Putra, H. W., & Arwani, I. (2021). Analisis sentimen pada media sosial Twitter terhadap aturan pemberlakuan pembatasan kegiatan masyarakat Kota Malang. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(10), 4486–4493.
- Nugroho, A. W., & Norhikmah. (2024). Analisis sentimen menggunakan algoritma Support Vector Machine pada Covid-19. *Jurnal Teknologi Informasi dan Komunikasi*, 13(4), 1758–1772.
- Putra, M. N. (2024). Bahasa pemrograman Java. *Jurnal Informatika dan Sistem Informasi*, 3(1), 3–5.
- Septiani, D., & Isabela, I. (2022). Analisis term frequency inverse document frequency (TF-IDF) dalam temu kembali informasi pada dokumen teks. *Jurnal Ilmu Komputer dan Matematika*, 25, 81–88.
- Styawati, Hendrastuty, N., Isnain, A. R., & Rahmadhani, A. Y. (2021). Analisis sentimen masyarakat terhadap program kartu prakerja pada Twitter dengan metode Support Vector Machine. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 6(3), 150–155.