Sistemasi: Jurnal Sistem Informasi
Volume 15, Nomor 5, 2026: 1635-1643
ISSN:2302-8149
e-ISSN:2540-9719

Evaluasi Pengaruh Kualitas Pelabelan dan Ketidakseimbangan Kelas pada Klasifikasi Sentimen Konflik Palestina-Israel

Evaluation of the Impact of Labeling Quality and Class Imbalance on Sentiment Classification of the Palestine–Israel Conflict

[1]Salvia Devi Muhshanah, [2]Evi Maria*
[1,2]Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
[1,2]Jl. Dr. O. Notohamidjodjo Blotongan, Sidorejo, Kota Salatiga, Jawa Tengah, Indonesia
*e-mail: evi.maria@uksw.edu

(***received*:** 9 April 2026, ***revised*:** 6 May 2026, ***accepted*:** 21 May 2026)

Abstrak

Penelitian ini bertujuan untuk mengevaluasi kinerja klasifikasi sentimen pada data media sosial terkait konflik Palestina-Israel dengan menekankan pada peran kualitas pelabelan dan distribusi data. Pendekatan yang digunakan mengombinasikan representasi teks TF-IDF dengan pelabelan berbasis *lexicon-based* (InSet) serta dua algoritma klasifikasi, yaitu Support Vector Machine (SVM) dan Random Forest. Data diperoleh dari platform X sebanyak 2.831 tweet berbahasa Indonesia yang telah melalui proses *preprocessing*. Hasil penelitian menunjukkan bahwa distribusi sentimen didominasi oleh kelas negatif (39,35%), diikuti netral (38,43%) dan positif (22,21%), yang mengindikasikan *imbalance* kelas. Evaluasi validitas pelabelan menunjukkan nilai Cohen's Kappa sebesar 0,0175, yang mengindikasikan rendahnya kesesuaian antara pelabelan otomatis dan anotasi manual. Kinerja model SVM memperoleh akurasi sebesar 0,69 dan *weighted F1-score* 0,68. Namun, kedua model menunjukkan performa rendah pada kelas positif sebagai kelas minoritas. Temuan ini menunjukkan bahwa keterbatasan performa model tidak semata-mata disebabkan oleh algoritma, melainkan dipengaruhi secara signifikan oleh kualitas pelabelan dan karakteristik distribusi data. Penelitian ini memberikan kontribusi dalam menekankan pentingnya evaluasi pipeline analisis sentimen secara menyeluruh, khususnya data yang kompleks dan tidak terkontrol seperti media sosial.

Kata kunci: imbalance data, *lexicon-based modeling*, analisis sentimen, TF-IDF, klasifikasi teks

Abstract

This study aims to evaluate the performance of sentiment classification on social media data related to the Palestine–Israel conflict, with a particular emphasis on the role of labeling quality and data distribution. The proposed approach combines TF-IDF text representation with lexicon-based labeling using InSet, along with two classification algorithms: Support Vector Machine (SVM) and Random Forest. The dataset was collected from the social media platform X and consisted of 2,831 Indonesian-language tweets that had undergone preprocessing. The results indicate that the sentiment distribution was dominated by the negative class (39.35%), followed by neutral (38.43%) and positive (22.21%) classes, indicating the presence of class imbalance. The labeling validity evaluation produced a Cohen's Kappa value of 0.0175, indicating a low level of agreement between automatic labeling and manual annotation. The SVM model achieved an accuracy of 0.69 and a weighted F1-score of 0.68. However, both models demonstrated poor performance on the positive class as the minority class. These findings suggest that the limitations in model performance are not solely caused by the classification algorithms themselves, but are also significantly influenced by labeling quality and data distribution characteristics. This study contributes by emphasizing the importance of comprehensive evaluation throughout the sentiment analysis pipeline, particularly when dealing with complex and uncontrolled data sources such as social media.

Keywords: *class imbalance, lexicon-based labeling, sentiment analysis, TF-IDF, text classification*

 ISSN:2302-8149
e-ISSN:2540-9719

1 Pendahuluan

Perkembangan teknologi digital telah mengubah cara masyarakat mengekspresikan opini publik, dengan media sosial seperti X (Twitter) menjadi ruang utama diskursus global yang berlangsung secara real-time dan terbuka. Dalam konteks ini, analisis teks (*text mining*) menjadi pendekatan penting untuk mengubah data tidak terstruktur menjadi informasi yang bermakna, khususnya dalam memahami dinamika opini publik terhadap isu-isu global yang kompleks [1][2]. Salah satu isu yang memicu intensitas diskursus global yang tinggi adalah konflik Palestina-Israel, yang tidak hanya berdampak secara geopolitik [3], tetapi juga menghasilkan respons emosional yang kuat dalam ruang digital. Media sosial menjadi medium utama dalam merepresentasikan spektrum sentimen publik, mulai dari ekspresi simpati, kecaman, hingga penyampaian informasi faktual. Karakteristik ini menjadikan data media sosial bersifat kompleks, tidak terstruktur, serta mengandung ambiguitas linguistik dan *noise* yang tinggi [4][5].

Namun demikian, analisis sentimen terhadap data media sosial dalam berbagai studi sebelumnya menunjukkan sejumlah tantangan metodologis yang signifikan. Literatur mengindikasikan bahwa data media sosial sering kali memiliki distribusi kelas yang tidak seimbang (*class imbalance*), yang dapat menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas [6]. Selain itu, kompleksitas bahasa dalam media sosial juga menyulitkan proses representasi teks secara akurat, terutama pada pendekatan berbasis *bag-of-words* seperti TF-IDF yang memiliki keterbatasan dalam menangkap konteks semantik yang lebih dalam [7]. Tantangan-tantangan ini menunjukkan bahwa performa model klasifikasi tidak hanya ditentukan oleh algoritma, tetapi juga oleh karakteristik data yang digunakan.

Pendekatan pelabelan berbasis *lexicon-based* merupakan salah satu strategi yang umum digunakan dalam analisis sentimen, terutama ketika anotasi manual dalam skala besar tidak tersedia. Pendekatan ini memungkinkan proses pelabelan awal secara efisien dan konsisten dalam praktik pengolahan data teks [8][9]. Sejumlah penelitian sebelumnya telah menerapkan metode klasifikasi seperti Support Vector Machine (SVM) dan Random Forest dalam analisis sentimen media sosial, termasuk pada isu konflik internasional [10][11][12]. SVM dikenal efektif dalam menangani data berdimensi tinggi, sementara Random Forest memiliki keunggulan dalam mengelola variabilitas data dan *noise* [13][14]. Namun, sebagian besar studi sebelumnya berfokus pada peningkatan akurasi model dan perbandingan algoritma, tanpa secara eksplisit mengevaluasi bagaimana kualitas pelabelan awal dan distribusi data memengaruhi performa model dalam pipeline yang merepresentasikan praktik nyata. Akibatnya, masih terbatas bukti empiris yang secara langsung mengevaluasi performa pendekatan umum dalam *pipeline* analisis sentimen yang merepresentasikan kondisi data nyata.

Berdasarkan celah tersebut, penelitian ini bertujuan untuk mengevaluasi performa klasifikasi sentimen menggunakan kombinasi TF-IDF dan pelabelan berbasis *lexicon-based*, dengan membandingkan algoritma Support Vector Machine (SVM) dan Random Forest. Pendekatan ini dipilih karena merepresentasikan praktik umum dalam analisis sentimen, sehingga memungkinkan evaluasi performa model dalam kondisi yang mendekati implementasi nyata. Fokus utama penelitian ini bukan pada perbandingan keunggulan algoritma, melainkan bagaimana karakteristik data, khususnya kualitas pelabelan dan distribusi kelas memengaruhi hasil klasifikasi dalam kondisi yang tidak sepenuhnya dapat dikendalikan secara eksperimental. Secara khusus, penelitian ini memberikan dua kontribusi utama. Pertama, memberikan evaluasi empiris terhadap performa model klasifikasi dalam konteks data yang memiliki karakteristik kompleks sebagaimana umum ditemukan pada media sosial. Kedua, menunjukkan bahwa keterbatasan performa model tidak selalu berasal dari algoritma yang digunakan, tetapi juga dari karakteristik data dan pendekatan pelabelan yang diterapkan. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih kritis terhadap penggunaan metode klasifikasi dalam analisis sentimen, serta menjadi dasar bagi pengembangan pendekatan yang lebih kontekstual dan *robust* dalam mengolah data opini publik dan isu-isu sosial yang kompleks.

2 Tinjauan Literatur

Penelitian mengenai analisis sentimen pada isu konflik internasional telah banyak menggunakan pendekatan klasifikasi berbasis *machine learning*, khususnya Support Vector Machine (SVM) dan Random Forest, dengan fokus utama pada peningkatan akurasi model. Beberapa studi melaporkan performa tinggi, seperti *F1-Score* mencapai 98% pada data komentar YouTube [10] serta akurasi di atas 79% pada *platform* X [11]. Studi lain juga menunjukkan bahwa pendekatan berbasis TF-IDF

 ISSN:2302-8149
e-ISSN:2540-9719

dengan SVM mampu menghasilkan akurasi di atas 90% pada klasifikasi sentimen [12]. Meskipun demikian, capaian performa tersebut umumnya diperoleh dalam konteks yang terbatas, baik dari sisi *platform*, skema klasifikasi (misalnya biner), maupun pendekatan pelabelan yang digunakan. Sebagian besar penelitian masih mengasumsikan label yang digunakan telah merepresentasikan sentimen secara valid, tanpa secara eksplisit mengevaluasi kualitas pelabelan atau potensi bias yang dihasilkan oleh metode pelabelan otomatis.

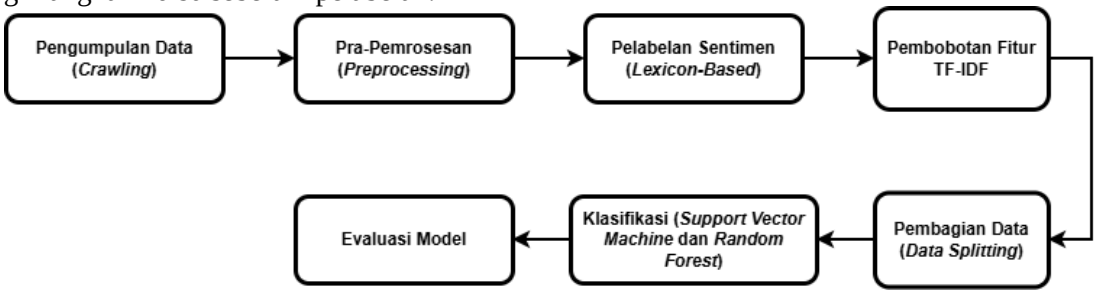
Di sisi lain, literatur menunjukkan bahwa pendekatan *lexicon-based*, yang sering digunakan dalam kondisi keterbatasan data berlabel, memiliki keterbatasan dalam menangkap makna kontekstual karena bergantung pada asosiasi kata yang bersifat statis [8]. Keterbatasan ini berpotensi menghasilkan label yang tidak sepenuhnya mencerminkan makna aktual dalam teks, terutama pada domain yang kompleks seperti konflik, di mana kata yang sama dapat digunakan dalam konteks informatif maupun evaluatif. Kondisi ini berkaitan dengan fenomena label *noise* yang telah diidentifikasi sebagai faktor yang dapat menurunkan performa model klasifikasi [9]. Selain itu, karakteristik data media sosial yang cenderung tidak seimbang juga menjadi tantangan dalam proses klasifikasi, karena model lebih mudah menggeneralisasi kelas mayoritas dibandingkan kelas minoritas [6]. Namun, sebagian besar penelitian sebelumnya lebih menekankan pada perbandingan performa algoritma, tanpa mengkaji secara sistematis bagaimana interaksi antara distribusi data, kualitas pelabelan, dan metode representasi teks memengaruhi hasil klasifikasi secara keseluruhan.

Dengan demikian, meskipun berbagai pendekatan telah dikembangkan dalam analisis sentimen, masih terdapat keterbatasan dalam memahami bagaimana *pipeline* analisis yang umum digunakan, yang menggabungkan TF-IDF, pelabelan berbasis *lexicon-based*, dan algoritma klasifikasi berperforma dalam kondisi data yang tidak sepenuhnya terkontrol. Kekosongan ini menunjukkan bahwa evaluasi terhadap kualitas data dan proses pelabelan menjadi aspek krusial, serta menegaskan pentingnya pendekatan yang berfokus pada evaluasi pipeline analisis sentimen secara keseluruhan, bukan semata-mata pada optimalisasi algoritma.

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis analisis teks untuk mengklasifikasikan sentimen opini publik terhadap konflik Palestina-Israel pada *platform* X. Klasifikasi sentimen dilakukan ke dalam tiga kategori, yaitu positif, negatif, dan netral, dengan tujuan mengkaji kinerja pelabelan sentimen berbasis *lexicon-based* serta implikasinya terhadap hasil klasifikasi pada konteks data yang berpotensi tidak seimbang dan sensitif secara semantik.

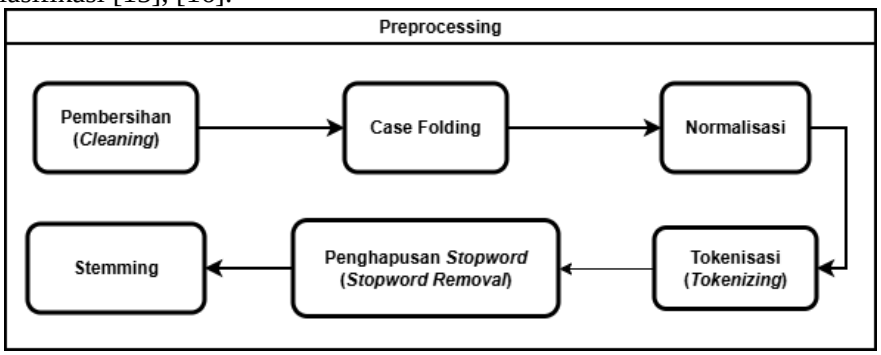
Alur penelitian disajikan pada Gambar 1, yang meliputi tahap pengumpulan data (*crawling*), pra-pemrosesan (*Preprocessing*), pelabelan sentimen berbasis *lexicon-based*, representasi fitur menggunakan TF-IDF, pembagian data, proses klasifikasi menggunakan Support Vector Machine (SVM) dan Random Forest, serta evaluasi kinerja model. Data dikumpulkan dari *platform* X menggunakan *library snscrape* (Python) dengan teknik *purposive sampling* pada periode Oktober 2023 hingga Desember 2025 menggunakan kata kunci “palestine”, “gaza”, “genosida”, dan “zionis”. Pemilihan kata kunci didasarkan pada istilah dominan digunakan dalam diskursus publik terkait konflik tersebut di media sosial. Data dibatasi pada *tweet* berbahasa Indonesia dan hanya mencakup *tweet* asli (*non-retweet*) untuk menghindari duplikasi konten. Dari proses *crawling* diperoleh 3.000 tweet, yang kemudian disaring menjadi 2.897 data unik setelah penghapusan duplikasi dan *spam* menggunakan Pandas 2.0. Data hasil *crawling* kemudian diproses melalui tahap *preprocessing* untuk menghilangkan *noise* sebelum pelabelan.



Gambar 1 Alur penelitian

 ISSN:2302-8149
e-ISSN:2540-9719

Tahap *preprocessing* disajikan pada Gambar 2 dan dilakukan untuk meningkatkan kualitas data melalui mengurangi *noise* serta standarisasi teks sebelum proses analisis [10]. Secara operasional, proses pembersihan (*cleaning*) dilakukan dengan menghapus URL, mention (@), hashtag (#), angka, tanda baca, serta emotikon menggunakan ekspresi reguler, kemudian menggantinya dengan spasi untuk menjaga struktur token. Selanjutnya, *case folding* dilakukan dengan mengubah seluruh teks menjadi huruf kecil. Normalisasi diterapkan untuk mengonversi kata tidak baku, singkatan, dan bentuk kasual ke bentuk standar menggunakan kamus normalisasi sederhana (misalnya "bbrp" menjadi "beberapa"). Setelah itu, teks diubah menjadi token kata melalui proses tokenisasi (*tokenizing*), kemudian dilakukan penghapusan *stopword* menggunakan daftar *stopword* bahasa Indonesia dari NLTK. Tahap akhir adalah *stemming* menggunakan pustaka Sastrawi untuk mengembalikan kata ke bentuk dasar dengan menghilangkan imbuhan. Seluruh tahapan ini bertujuan untuk menghasilkan representasi teks yang lebih konsisten dan relevan sebagai input dalam proses pelabelan dan klasifikasi [15], [16].



Gambar 2 Tahapan *preprocessing*

Pelabelan sentimen dilakukan menggunakan pendekatan *lexicon-based* dengan memanfaatkan InSet Lexicon yang dikembangkan untuk bahasa Indonesia oleh Koto dan Rahmaningtyas [17]. Pendekatan ini merupakan salah satu metode umum dalam analisis sentimen berbasis kamus [8]. *Lexicon-based* terdiri dari 3.609 kata positif dan 6.609 kata negatif dengan rentang bobot sentimen dari -5 hingga +5. Skor sentimen setiap tweet dihitung berdasarkan agregasi bobot polaritas kata yang terdapat dalam teks. Hasil akumulasi kemudian diklasifikasikan ke dalam tiga kategori, yaitu positif: skor $>$ 0,5; negatif: skor $<$ -0,5; netral: -0,5 $\leq$ skor $\leq$ 0,5. Pendekatan ini dipilih karena memungkinkan pelabelan otomatis dalam skala besar tanpa memerlukan anotasi manual yang memakan waktu dan biaya. Selain itu, metode *lexicon-based* banyak digunakan sebagai pendekatan awal (*baseline)* dalam analisis sentimen, terutama pada konteks data berbahasa Indonesia yang memiliki keterbatasan dataset berlabel. Namun demikian, *lexicon-based* yang digunakan bersifat umum dan tidak secara spesifik dirancang untuk domain konflik, sehingga berpotensi menghasilkan pelabelan yang tidak sepenuhnya menangkap makna kontekstual dalam teks.

Pelabelan ini dikategorikan sebagai *weak supervision* karena tidak menggunakan anotasi manual sebagai *ground truth*. Oleh karena itu, dilakukan validasi terhadap 100 *tweet* yang dipilih secara acak dan dilabeli secara manual. Anotasi manual dilakukan oleh satu peneliti dengan pemahaman terhadap konteks isu konflik. Anotasi manual dilakukan berdasarkan pedoman berikut: (1) label positif diberikan *tweet* yang mengandung dukungan, empati, atau sentimen positif terhadap pihak tertentu; (2) label negatif diberikan pada *tweet* yang mengandung kritik, kecaman, atau ekspresi emosi negatif; dan (3) label netral diberikan pada *tweet* yang bersifat informatif atau tidak menunjukkan kecenderungan sentimen yang jelas. Tingkat kesesuaian antara pelabelan otomatis dan manual diukur menggunakan *agreement* dan Koefisien Cohen's Kappa (κ) untuk memperhitungkan peluang kesepakatan acak [18][19]. Nilai Kappa digunakan untuk mengevaluasi reliabilitas pelabelan, dengan interpretasi bahwa nilai κ di bawah 0,40 menunjukkan tingkat kesepakatan yang rendah dan mengindikasikan adanya perbedaan sistematis antara pendekatan *lexicon-based* dan interpretasi manusia [19]. Namun demikian, karena anotasi manual dilakukan oleh satu peneliti, evaluasi ini tidak merepresentasikan *inter-annotator agreement*, melainkan perbandingan antara pelabelan otomatis dan anotasi tunggal (*intra-rater comparison*). Oleh karena itu, interpretasi nilai Kappa dalam penelitian ini bersifat terbatas dan lebih diarahkan untuk mengidentifikasi potensi bias dalam pendekatan berbasis

 ISSN:2302-8149
e-ISSN:2540-9719

lexicon-based. Oleh karena itu, hasil validasi ini tidak dimaksudkan untuk mengukur reliabilitas anotasi secara umum, melainkan untuk mengidentifikasi potensi bias dalam pelabelan otomatis yang digunakan dalam *pipeline* analisis.

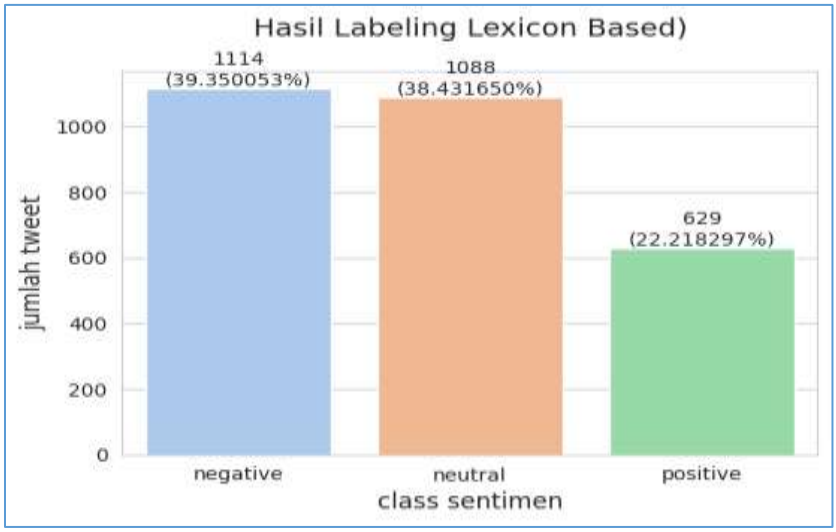
Representasi fitur dilakukan menggunakan metode TF-IDF untuk mengubah teks menjadi vektor numerik berdasarkan frekuensi kemunculan kata dalam dokumen dan distribusinya pada seluruh korpus [18]. Implementasi TF-IDF menggunakan parameter default dari scikit-learn dengan *ngram_range* = (1,1) dan tanpa pembatasan jumlah fitur (*max_features* = *None*). Dataset kemudian dibagi menjadi data latih dan data uji menggunakan fungsi *train_test_split* dari scikit-learn dengan rasio 80:20 dan parameter *random_state* = 42 untuk memastikan reproduksibilitas hasil.

Proses klasifikasi dilakukan menggunakan dua algoritma, yaitu Support Vector Machine (SVM) dengan kernel linear dan parameter regularisasi C = 1.0, serta Random Forest sebagai metode pembanding berbasis *ensemble* dengan jumlah pohon sebanyak 100 (*n_estimators* = *100*) dan *random_state* = 42 untuk memastikan konsistensi hasil. Pemilihan kernel linear pada SVM didasarkan pada karakteristik data teks berdimensi tinggi yang cenderung *sparse*, sehingga lebih efisien dan stabil dalam proses klasifikasi [11][14], sementara Random Forest digunakan untuk menguji *robustness* model terhadap variasi fitur dan *noise* pada data [20][21][22]. Evaluasi kinerja model dilakukan menggunakan metrik akurasi dan *weighted F1-score* untuk mempertimbangkan potensi ketidakseimbangan distribusi kelas. Seluruh proses analisis diimplementasikan menggunakan Python 3.10 dengan pustaka utama scikit-learn, Pandas 2.0, NumPy 1.24, NLTK, dan Sastrawi.

4 Hasil dan Pembahasan

Hasil Penelitian

Dataset yang dianalisis sebanyak 2.831 tweet setelah proses pembersihan data, dengan 66 data dieliminasi karena duplikasi dan data kosong. Distribusi sentimen berdasarkan pelabelan menggunakan pendekatan *lexicon-based* (Gambar 3) menunjukkan bahwa sentimen negatif mendominasi dengan 1.114 data (39,35%), diikuti oleh sentimen netral sebanyak 1.088 data (38,43%), dan sentimen positif sebanyak 629 data (22,21%). Distribusi ini menunjukkan adanya ketidakseimbangan kelas, terutama pada kelas positif yang memiliki proporsi lebih rendah dibandingkan kelas lainnya. Kondisi ini berpotensi memengaruhi kinerja model klasifikasi, khususnya dalam mengidentifikasi kelas minoritas secara akurat.



Gambar 3 Hasil *labeling lexicon-based*
Sumber: Data sekunder diolah, 2026

Evaluasi kualitas pelabelan dilakukan melalui uji validitas terhadap 100 sampel data. Hasil pengujian menunjukkan nilai agreement sebesar 0,38 dan Cohen's Kappa sebesar 0,0175 (Tabel 1), yang mencerminkan tingkat kesesuaian yang rendah antara pelabelan otomatis dan pelabelan manual. Pada tahap ekstraksi fitur menggunakan TF-IDF, diperoleh sebanyak 13.549 fitur kata unik. Dataset kemudian dibagi menjadi 2.264 data latih (80%) dan 567 data uji (20%). Berdasarkan pembobotan

 ISSN:2302-8149
e-ISSN:2540-9719

TF-IDF, lima *term* dengan skor tertinggi yang merepresentasikan topik dominan dalam dataset disajikan pada Tabel 2.

Tabel 1 Hasil uji validitas

Uji Validitas	Hasil
Agreement	0,38
Koefisien Cohen's Kappa	0,0175

Sumber: Data sekunder diolah (2026)

Tabel 2 Lima *term* dengan skor TF-IDF tertinggi

Term	Skor TF-IDF
Gaza	97,70
Genosida	92,09
Palestina	86,08
Israel	84,86
Zionis	84,52

Sumber: Data sekunder diolah (2026)

Hasil klasifikasi menunjukkan bahwa model Support Vector Machine (SVM) menghasilkan akurasi sebesar 0,69 dengan *weighted F1-score* sebesar 0,68, dengan Random Forest menghasilkan akurasi sebesar 0,66 dengan *weighted F1-score* sebesar 0,64 (Tabel 3). Secara lebih rinci, performa masing-masing model pada setiap kelas ditunjukkan pada Tabel 4. Model SVM menunjukkan performa terbaik pada kelas negatif dengan nilai *recall* sebesar 0,78, serta performa relatif seimbang pada kelas netral. Namun, performa pada kelas positif cenderung lebih rendah dengan nilai *recall* sebesar 0,40.

Tabel 3 Ringkasan Performa model

Model	*Accuracy*	*Weighted F1-score*
SVM	0,69	0,68
Random Forest	0,66	0,64

Sumber: Data sekunder diolah (2026)

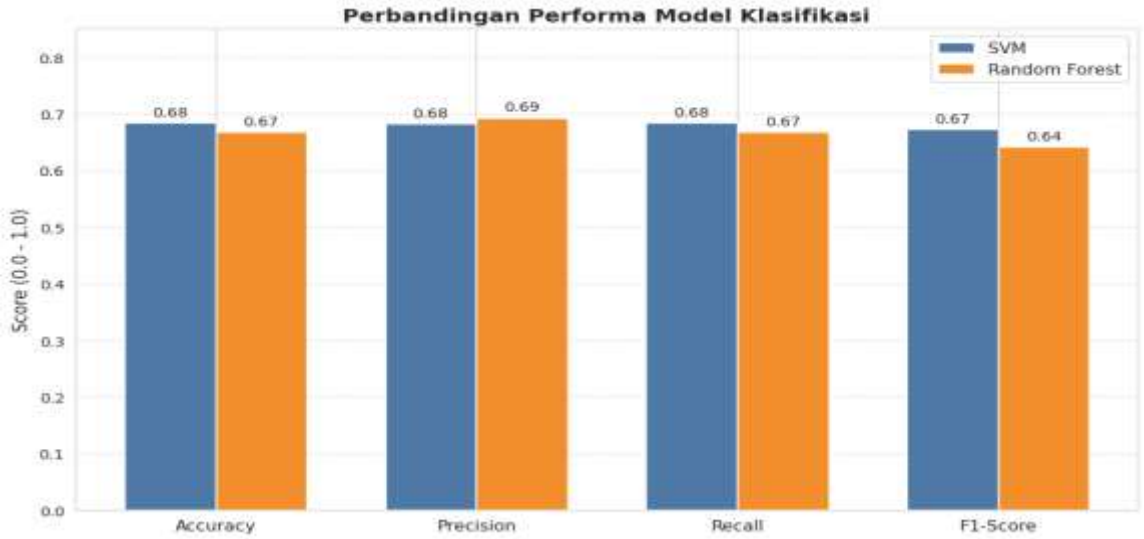
Tabel 4 Perbandingan hasil klasifikasi SVM dan *random forest*

Model	Kelas	*Precision*	*Recall*	*F1-score*
SVM	Negatif	0,73	0,78	0,75
	Netral	0,64	0,76	0,70
	Positif	0,76	0,40	0,52
Random Forest	Negatif	0,67	0,77	0,71
	Netral	0,64	0,79	0,71
	Positif	0,74	0,24	0,36

Sumber: Data sekunder diolah (2026)

Sebagai pembanding, Random Forest menunjukkan pola yang serupa, namun dengan penurunan performa yang lebih signifikan pada kelas positif, dengan nilai recall hanya sebesar 0,24. Hal ini mengindikasikan bahwa kedua model mengalami kesulitan dalam mengidentifikasi kelas positif secara akurat, terutama pada kondisi distribusi data yang tidak seimbang. Perbandingan kedua model disajikan pada Gambar 4 menunjukkan bahwa SVM memiliki performa yang sedikit lebih tinggi dibandingkan Random Forest, dengan selisih yang relatif terbatas pada metrik evaluasi yang digunakan.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 4 Perbandingan performa SVM dan *random forest*
Sumber: Data diolah (2026)

Diskusi

Hasil penelitian menunjukkan bahwa distribusi sentimen didominasi oleh kelas negatif, dengan proporsi yang jauh lebih tinggi dibandingkan sentimen positif. Temuan ini tidak hanya mencerminkan karakteristik data, tetapi juga menunjukkan bahwa analisis sentimen pada isu konflik bersenjata sangat dipengaruhi oleh konteks emosional dan narasi kemanusiaan yang berkembang di ruang publik digital. Studi sebelumnya menunjukkan bahwa diskursus konflik dan kekerasan cenderung memicu ekspresi afektif yang kuat, sehingga meningkatkan dominasi sentimen negatif dalam media sosial [1]. Dengan demikian, ketidakseimbangan kelas yang ditemukan dalam penelitian ini merupakan pola yang konsisten dengan karakteristik diskursus konflik dalam media sosial.

Selain faktor kontekstual, distribusi sentimen tersebut juga dipengaruhi oleh mekanisme pelabelan yang digunakan. Pendekatan *lexicon-based* cenderung memberikan polaritas negatif pada kata-kata yang secara leksikal berasosiasi dengan konflik, seperti perang, korban, dan serangan. Akibatnya, teks yang bersifat informatif atau deskriptif berpotensi terklasifikasi sebagai sentimen negatif. Hal ini menunjukkan bahwa ketidakseimbangan distribusi tidak sepenuhnya merefleksikan opini publik, tetapi juga dipengaruhi oleh bias sistematis dalam pendekatan *lexicon-based*. Keterbatasan ini dikonfirmasi oleh hasil uji validitas yang menunjukkan nilai Cohen's Kappa yang mendekati nol, mengindikasikan bahwa pelabelan berbasis *lexicon-based* dalam konteks ini tidak dapat dianggap sebagai representasi *ground truth*, melainkan sebagai pendekatan awal yang mengandung bias sistematis. Oleh karena itu, hasil validasi ini tidak dimaksudkan untuk mengukur reliabilitas anotasi secara umum, melainkan untuk mengidentifikasi potensi bias dalam pelabelan otomatis.

Temuan ini memperlihatkan keterbatasan pendekatan *lexicon-based* dalam menangkap makna kontekstual, khususnya pada domain konflik. *Lexicon sentiment analysis* pada dasarnya bergantung pada asosiasi kata yang statis, sehingga tidak mampu membedakan antara penggunaan kata dalam konteks ekspresi opini, laporan informasi, maupun ironi [8][4][2]. Dalam konteks penelitian ini, kata-kata seperti "genosida" atau "serangan" secara sistematis diberi polaritas negatif oleh sistem, meskipun dalam penggunaan aktual dapat muncul dalam konteks informatif yang netral. Hal ini menjelaskan rendahnya kesesuaian antara pelabelan otomatis dan manual, sekaligus menunjukkan bahwa pendekatan *lexicon-based* kurang memadai untuk analisis sentimen pada isu kompleks dan sensitif.

Keterbatasan kualitas label ini berimplikasi langsung pada performa model klasifikasi. Meskipun model *Support Vector Machine* menunjukkan performa yang sedikit lebih tinggi dibandingkan *Random Forest*, kedua model sama-sama mengalami penurunan kinerja pada kelas positif. *Error analysis* menunjukkan bahwa kesalahan klasifikasi terutama pada kelas positif sebagai kelas minoritas, yang sering diklasifikasikan sebagai netral atau negatif, mengindikasikan adanya bias model terhadap kelas mayoritas. Fenomena ini dapat dijelaskan dengan dua faktor utama. Pertama, ketidakseimbangan distribusi kelas menyebabkan model cenderung bias terhadap kelas mayoritas [6].

 ISSN:2302-8149
e-ISSN:2540-9719

Kedua, *noise* pada label akibat ketidaksesuaian *lexicon-based* memperburuk proses pembelajaran model, karena algoritma dilatih menggunakan representasi yang tidak sepenuhnya mencerminkan makna yang sebenarnya [9]. Oleh karena itu, performa model yang relatif moderat dalam penelitian ini lebih tepat dipahami sebagai refleksi dari kualitas data dan karakteristik domain, bukan semata-mata keterbatasan algoritma.

Selain itu, perbedaan performa antara SVM dan *Random Forest* yang relatif kecil menunjukkan bahwa pemilihan algoritma tampaknya bukan faktor yang paling menentukan dalam konteks ini. SVM yang berbasis margin optimal cenderung lebih stabil dalam menangani data berdimensi tinggi seperti TF-IDF, sehingga menghasilkan performa yang sedikit lebih baik. Namun, tanpa perbaikan pada kualitas pelabelan dan distribusi data, peningkatan performa model akan tetap terbatas. Temuan ini sejalan dengan penelitian Zhang et al. [7] yang menekankan bahwa klasifikasi teks, kualitas data seringkali lebih menentukan dibandingkan kompleksitas model.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa analisis sentimen pada isu konflik tidak dapat dijelaskan secara memadai hanya dengan pendekatan berbasis *lexicon-based* dan model klasifikasi konvensional, tetapi diperlukan pendekatan yang lebih kontekstual, seperti *supervised learning* dengan anotasi manual yang lebih reliabel atau penggunaan model berbasis pembelajaran mendalam yang mampu menangkap konteks semantik secara lebih baik. Dengan demikian, kontribusi utama penelitian ini terletak pada pengungkapan bahwa keterbatasan performa model dalam analisis sentimen pada isu konflik lebih ditentukan oleh kualitas pelabelan dan kesesuaian representasi bahasa terhadap domain dibandingkan oleh algoritma semata.

5 Kesimpulan

Penelitian ini menunjukkan bahwa analisis sentimen terhadap konflik Palestina-Israel di *platform* X didominasi oleh sentimen negatif, dengan distribusi yang tidak seimbang antar kelas. Model *Support Vector Machine* menunjukkan performa yang sedikit lebih baik dibandingkan Random Forest, namun keduanya menghasilkan kinerja yang relatif moderat, terutama pada kelas positif. Hasil uji validitas pelabelan menunjukkan nilai Cohen's Kappa yang sangat rendah, yang mengindikasikan ketidaksesuaian antara pelabelan berbasis *lexicon-based* dan interpretasi manual. Temuan ini menegaskan bahwa kualitas pelabelan memiliki peran penting dalam menentukan performa model klasifikasi.

Penelitian ini memiliki keterbatasan pada penggunaan *lexicon-based* umum yang belum mampu menangkap konteks secara memadai serta validasi manual yang masih terbatas. Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan pendekatan pelabelan yang lebih kontekstual, seperti penggunaan *lexicon-based* spesifik domain atau anotasi manual dalam skala yang lebih luas. Selain itu, pemanfaatan model berbasis deep learning juga perlu dieksplorasi untuk meningkatkan kemampuan dalam memahami makna semantik pada isu-isu kompleks.

Referensi

[1] A. Giachanou and F. Crestani, "*Like it or not: A Survey of Twitter Sentiment Analysis Methods*," *ACM Comput. Surv.*, Vol. 49, No. 2, pp. 1–41, Jun. 2017, DOI: 10.1145/2938640.

[2] Y. Mao, Q. Liu, and Y. Zhang, "*Sentiment Analysis Methods, Applications, and Challenges: A Systematic Literature Review*," *J. King Saud Univ. - Comput. Inf. SCI.*, Vol. 36, No. 4, p. 102048, Apr. 2024, DOI: 10.1016/j.jksuci.2024.102048.

[3] R. Christie, G. Suha Ma'rifa, and J. A. Priliska, "Analisis Konflik Israel dan Palestina terhadap Pelanggaran Hak Asasi Manusia dalam Perspektif Hukum Internasional," *J. Kewarganegaraan*, Vol. 8, No. 1, pp. 349–350, 2024.

[4] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[5] S. M. Mohammad, "*Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text*," in *Emotion Measurement*, H. L. Meiselman, Ed., Elsevier, 2016, pp. 201–237. DOI: 10.1016/B978-0-08-100508-8.00009-6.

[6] Haibo He and E. A. Garcia, "*Learning from Imbalanced Data*," *IEEE Trans. Knowl. Data Eng.*, Vol. 21, No. 9, pp. 1263–1284, Sep. 2009, DOI: 10.1109/TKDE.2008.239.

[7] Y. Zhang, R. Jin, and Z.-H. Zhou, "*Understanding Bag-of-Words Model: A Statistical Framework*," *Int. J. Mach. Learn. Cybern.*, Vol. 1, No. 1–4, pp. 43–52, Dec. 2010, DOI:

 ISSN:2302-8149
e-ISSN:2540-9719

10.1007/s13042-010-0001-0.

[8] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "*Lexicon-based Methods for Sentiment Analysis*," *Comput. Linguist.*, Vol. 37, No. 2, pp. 267–307, Jun. 2011, DOI: 10.1162/COLI_a_00049.

[9] B. Frenay and M. Verleysen, "*Classification in the Presence of Label Noise: A Survey*," *IEEE Trans. Neural Networks Learn. Syst.*, Vol. 25, No. 5, pp. 845–869, May 2014, DOI: 10.1109/TNNLS.2013.2292894.

[10] A. A. Syam, G. Hardy M, A. Salim, D. F. Surianto, and M. Fajar B, "Analisis Teknik *Preprocessing* pada Sentimen Masyarakat terkait Konflik Israel-Palestina menggunakan *Support Vector Machine*," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, Vol. 9, No. 3, pp. 1465–1467, Aug. 2024, DOI: 10.29100/jipi.v9i3.5527.

[11] D. Deltania, Garno, and A. Jamaludin, "Analisis Sentimen Publik terhadap Invasi Zionis kepada HAMAS menggunakan *Support Vector Machine (SVM)*," *J. Mhs. Tek. Inform.*, Vol. 8, No. Vol. 8 No. 4 (2024): JATI Vol. 8 No. 4, pp. 4465–4466, Aug. 2024, DOI: https://doi.org/10.36040/jati.v8i4.9959.

[12] F. M. Carina, Admi Salma, Dony Permana, and Zamahsary Martha, "*Sentiment Analysis of X Application Users on the Conflict between Israel and Palestine using Support Vector Machine Algorithm*," *UNP J. Stat. Data SCI.*, Vol. 2, No. 2, p. 204, May 2024, DOI: 10.24036/ujsds/vol2-iss2/170.

[13] L. Breiman, "*Random Forests*," *Mach. Learn.*, Vol. 45, No. 1, pp. 5–32, Oct. 2001, DOI: 10.1023/A:1010933404324.

[14] T. Wahyudi *et al.*, "Klasifikasi Sentimen *X-Twitter* Perihal Pemindahan Ibu Kota Indonesia menggunakan Ekstrasi Fitur *TF-IDF* dan Metode S*upport Vector Machine (SVM)*," *J. Keilmuan dan Apl. Bid. Tek. Inform.*, Vol. 18, No. 2, p. 191, Aug. 2024, DOI: 10.47111/JTI.

[15] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna *Twitter* terhadap Layanan Internet Provider menggunakan Algoritma *Support Vector Machine*," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, Vol. 20, No. 2, p. 410, May 2021, DOI: 10.30812/matrik.v20i2.1130.

[16] R. Azhar and M. F. Wijayanto, "Analisis Sentimen di Twitter: Mengungkap Persepsi dan Emosi Publik Seputar Konflik Palestina-Israel," *Stain. (Seminar Nas. Teknol. Sains)*, Vol. 3, No. Vol. 3 No. 1 (2024): STAINS (Seminar Nasional Teknologi & Sains), p. 120, 2024, DOI: https://doi.org/10.29407/stains.v3i1.4132.

[17] F. Koto and G. Y. Rahmaningtyas, "*Inset Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs*," Singapura, 2017. DOI: 10.1109/IALP.2017.8300625.

[18] D. Ananda Efraim, "Analisis Sentimen pada Sosial Media Instagram menggunakan Algoritma *Naive Bayes* (Studi Kasus : Timnas Futsal Indonesia)," Jakarta, Aug. 2023.

[19] J. R. Landis and G. G. Koch, "*The Measurement of Observer Agreement for Categorical Data*," *Biometrics*, Vol. 33, No. 1, pp. 159–174, Mar. 1977, DOI: 10.2307/2529310.

[20] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi *Random Forest* dan *SVM* pada Analisis Sentimen PSBB," *J. Inform. UPGRIS*, Vol. 7, p. 39, Jun. 2021, Accessed: Mar. 29, 2026. [Online]. Available: https://journal.upgris.ac.id/index.php/JIU/article/view/7099/4309

[21] I. Afdhal *et al.*, "Penerapan Algoritma *Random Forest* untuk Analisis Sentimen Komentar di *YouTube* tentang Islamofobia," *J. Nas. Komputasi dan Teknol. Inf.*, Vol. 5, No. 1, p. 124, Feb. 2022, Accessed: Mar. 29, 2026. [Online]. Available: https://repository.uin-suska.ac.id/59747/1/Jurnal Ibnu Afdhal.pdf

[22] P. Kumala Sari and R. Randy Suryono, "Komparasi Algoritma *Support Vector Machine* dan *Random Forest* untuk Analisis *Sentimen Metaverse*," *J. Mnemon.*, Vol. 7, No. 1, pp. 32–33, Feb. 2024, Accessed: Mar. 29, 2026. [Online]. Available: https://www.ejournal.itn.ac.id/mnemonic/article/view/8977