RESEARCH ARTICLE • OPEN ACCESS

Penerapan Model Word Embedding Indobertweet dengan Metode Support Vector Machine dalam Klasifikasi Opini Publik

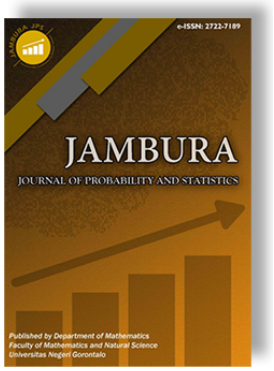
Naufal Daffa Faiz Pahrun, Novianita Achmad, and Isran K. Hasan

Volume 7, Issue 1, Pages 23–28, May 2026

Received 18 May 2025, Revised 20 May 2025, Accepted 20 Juni 2026, Published Online 9 Juli 2025

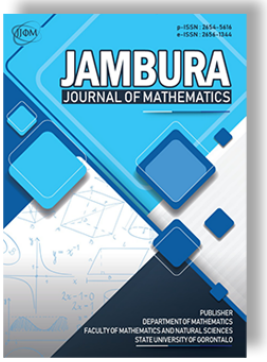
To Cite this Article : N. D. F. Pahrun, N. Achmad, and I, K, Hasan,“ Penerapan Model Word Embedding Indobertweet dengan Metode Support Vector Machine dalam Klasifikasi Opini Publik ”, *Jambura J. Probab. Stat.*, vol. 7, no. 1, pp. 23–28, 2026, *https://doi.org/10.34312/jjps.v7i1.38992*

JOURNAL INFO • JAMBURA JOURNAL OF PROBABILITY AND STATISTICS



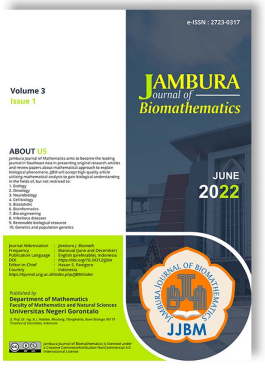
Homepage	:	https://ejurnal.ung.ac.id/index.php/jps/index
Journal Abbreviation	:	Jambura J. Probab. Stat.
Frequency	:	Biannual (May and November)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.34312/jjps
Online ISSN	:	2722-7189
Editor-in-Chief	:	Ismail Djakaria
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jps/oai
Google Scholar ID	:	kWdujzMAAAAJ
Email	:	redaksi.jjps@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

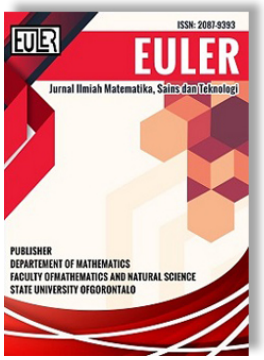


Jambura Journal of Mathematics

Jambura Journal of Mathematics Education



Jambura Journal of Biomathematics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi

Jambura Journal of Probability and Statistics, Volume 7, Issue 1, Pages 23–28, May 2026 https://doi.org/10.34312/jjps.v7i1.38992

Research Article

Check for updates

Penerapan Model Word Embedding Indobertweet dengan Metode Support Vector Machine dalam Klasifikasi Opini Publik

Naufal Daffa Faiz Pahrun[1*], **Novianita Achmad**[1], **Isran K. Hasan**[1]

[1,2,3] *Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Gorontalo*

ARTICLE HISTORY
Received 18 May 2025
Revised 20 May 2025
Accepted 20 Juni 2026
Published 9 Juli 2025

KATA KUNCI
Analisis Sentimen, SVM, IndoBERTweet, Media Sosial, Redenominasi Rupiah.

KEYWORDS
Sentiment Analysis, SVM, IndoBERTweet, Social Media, Rupiah Redenomination

ABSTRAK. *Penelitian ini bertujuan untuk mengklasifikasikan sentimen opini publik terhadap isu redenominasi rupiah pada media sosial X menggunakan kombinasi metode Support Vector Machine (SVM) dan word embedding IndoBERTweet. Permasalahan utama dalam analisis sentimen media sosial terletak pada karakteristik teks yang tidak terstruktur, penggunaan bahasa informal, serta adanya ambiguitas konteks. Oleh karena itu, diperlukan pendekatan yang mampu memahami konteks bahasa secara lebih mendalam sekaligus menghasilkan klasifikasi yang akurat. Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan pengumpulan data melalui teknik crawling, preprocessing teks, pelabelan data, ekstraksi fitur menggunakan IndoBERTweet, serta klasifikasi menggunakan SVM dengan kernel Radial Basis Function (RBF). Data yang digunakan sebanyak 1.014 tweet, kemudian diseleksi menjadi 370 data yang terdiri dari sentimen positif dan negatif. Hasil penelitian menunjukkan bahwa model yang diusulkan mampu mencapai akurasi sebesar 82%, precision 83%, recall 82%, dan F1-score 82%. Temuan ini menunjukkan bahwa kombinasi IndoBERTweet dan SVM efektif dalam menangkap makna kontekstual bahasa Indonesia, khususnya pada teks media sosial, serta mampu meningkatkan kinerja klasifikasi sentimen. Selain itu, hasil analisis menunjukkan bahwa mayoritas opini publik cenderung positif terhadap isu redenominasi rupiah. Penelitian ini diharapkan dapat menjadi referensi dalam pengembangan analisis sentimen berbasis machine learning serta mendukung pengambilan kebijakan yang lebih responsif terhadap opini masyarakat.*

ABSTRACT. *This study aims to classify public sentiment regarding the redenomination of the Indonesian rupiah on social media platform X using a combination of Support Vector Machine (SVM) and IndoBERTweet word embedding. The main challenge in social media sentiment analysis lies in unstructured text, informal language usage, and contextual ambiguity. Therefore, an approach capable of capturing contextual meaning while maintaining high classification accuracy is required. This research employs a quantitative approach, including data collection through crawling, text preprocessing, data labeling, feature extraction using IndoBERTweet, and classification using SVM with a Radial Basis Function (RBF) kernel. A total of 1,014 tweets were collected and refined into 370 labeled data consisting of positive and negative sentiments. The results show that the proposed model achieves an accuracy of 82%, precision of 83%, recall of 82%, and F1-score of 82%. These findings indicate that the integration of IndoBERTweet and SVM effectively captures contextual semantics in Indonesian social media text and improves sentiment classification performance. Furthermore, the analysis reveals that the majority of public opinions tend to be positive toward the redenomination issue. This study is expected to contribute to the development of machine learning-based sentiment analysis and support more responsive policy-making based on public opinion*

1. Pendahuluan

Perkembangan teknologi informasi, khususnya media sosial, telah mengubah cara masyarakat membentuk dan menyampaikan opini publik. Media sosial seperti X memungkinkan pengguna mengekspresikan pandangan terhadap berbagai isu strategis secara cepat dan luas, sehingga menjadi sumber data penting dalam memahami persepsi masyarakat terhadap kebijakan publik [1]. Selain itu, media sosial juga berperan sebagai ruang diskusi publik yang dinamis dan mampu mempengaruhi pembentukan opini kolektif [2]. Dalam konteks ini, isu redenominasi rupiah menjadi salah satu topik yang banyak diperbincangkan karena berkaitan dengan stabilitas ekonomi dan kepercayaan masyarakat terhadap mata uang nasional [3]. Permasalahan utama dalam penelitian ini terletak pada kompleksitas analisis opini publik di media sosial yang cenderung tidak terstruktur dan mengandung bahasa informal. Teks media sosial sering kali bersifat ambigu dan menggunakan bahasa tidak baku, sehingga menyulitkan proses klasifikasi sentimen secara akurat [4]. Selain itu dinamika opini masyarakat terhadap kebijakan redenominasi menunjukkan adanya perbedaan presepsi yang belum terpetakan secara objektif berbasis data [5]. Kondisi ini menuntut adanya pendekatan analisis yang mampu menangkap kompleksitas bahasa sekaligus menghasilkan klasifikasi yang andal

*Corresponding Author.

Email : *daffapahrun@gmail.com* (N. D. F. Pahrun)
Homepage : https://ejurnal.ung.ac.id/index.php/jps/index / E-ISSN : 2722-7189

[6]. Sebagai solusi, penelitian ini mengusulkan pendekatan analisis sentimen berbasis kombinasi *support vector machine* (SVM) dan *word embedding* kontekstual IndoBERTweet yang mampu menghasilkan pemisahan kelas optimal dengan tingkat generalisasi yang baik [7]. Metode ini juga terbukti efektif dalam berbagai studi klasifikasi teks dengan tingkat akurasi yang tinggi [8]. Untuk meningkatkan pemahaman konteks bahasa, digunakan IndoBERTweet yang mampu merepresentasikan makna kata secara kontekstual, khususnya pada teks informal media sosial [9]. Pendekatan ini diperkuat dengan teknik *word embedding* yang dapat menangkap hubungan semantik antar kata secara lebih komperhensif [10]. Penelitian-penelitian sebelumnya menunjukkan bahwa SVM memiliki performa lebih baik dibandingkan metode lain seperti KNN dan *Naive Bayes* dalam klasifikasi teks [11]. Disisi lain, model berbasis BERT terbukti mampu meningkatkan akurasi analisis sentimen karena mempertimbangkan konteks dua arah dalam pemrosesan bahasa alami [9]. Namun, sebagian besar penelitian masih berfokus pada deteksi ujaran kebencian dan emosi, sehingga penerapan IndoBERTweet sebagai ekstraksi fitur dalam analisis sentimen isu kebijakan ekonomi masih terbatas [12]. Hal ini menunjukkan adanya celah peneitian yang menjadi dasar kebaruan studi ini. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengklasifikasikan sentimen opini publik di media sosial X terkait redenominasi rupiah menggunakan kombinasi metode SVM dan IndoBERTweet. Pendekatan ini diharapkan mampu menghasilkan analisis yang lebih akurat dan kontekstual dalam memahami presepsi masyarakat [5]. Selain itu, hasil penelitian ini diharapkan dapat memberikan kontribusi ilmiah sekaligus menjadi referensi dalam perumusan kebijakan yang lebih responsif terhadap opini publik [9].

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode machine learning untuk mengklasifikasikan sentimen opini publik terhadap isu redenominasi rupiah di media sosial X. Adapun penelitian ini mencakup tahapan berikut:

2.1. Pengambilan Data

Pengambilan data diperoleh dari unggahan komentar pengguna pada media sosial X menggunakan Teknik *crawling* dengan memanfaatkan *tool tweet harvest* dengan kata kunci "Redenominasi Rupiah" dalam rentang waktu 10 Oktober 2025 – 10 Januari 2026. Pengambilan data menghasilkan 1.014 data *tweet*.

2.2. Pelabelan Data

Pelabelan data ini bertujuan untuk mengkategorikan sentimen ke dalam dua kategori positif dan negatif yang masing□masing diberi label 1 untuk label positif dan □1 untuk label negatif. Sentimen positif diberikan pada tweet yang mendukung redenominasi rupiah, sedangkan sentimen negatif diberikan pada *tweet* yang menolak.

2.3. Preprocessing data

Preprocessing data merupakan hal krusial dalam analisis sentimen yang bertujuan untuk menyeragamkan data teks sebelum diproses pada metode klasifikasi. Tahapan *preprocessing* data meliputi tahapan berikut [13]:

1. *Case folding*, yaitu mengubah seluruh huruf dalam teks menjadi huruf kecil.
2. *Cleaning*, yakni membersihkan teks dari karakter yang tidak diperlukan seperti simbol diluar alfabet, tanda baca, tautan/URL, *hashtag* dan *username*.
3. *Tokenizing*, yaitu memecah kalimat menjadi unit-unit kata berdasarkan struktur penyusunnya.
4. *Normalization*, yaitu mengubah kata tidak baku menjadi bentuk standar sesuai kaidah kebahasaan.
5. *Stemming*, yaitu mengubah kata berimbuhan bentuk dasarnya untuk mengurangi jumlah variasi kata yang perlu diproses oleh model.

2.4. Ekstraksi Fitur

Dalam penelitian ini, ekstraksi fitur dilakukan dengan menggunakan model IndoBERTweet, yaitu model bahasa pralatih berbasis arsitektur *Bidirectional Encoder Representations from Transofrmer* (BERT) yang dikembangkan khusus untuk teks media sosial berbahasa indonesia. IndoBERTweet merupakan penggabungan dari IndoBERT dengan penambahan kosakata yang relevan dengan karakteristik bahasa pada media sosial X, sehingga mampu memahami konteks bahasa informal, singkatan, dan ekspresi media sosial [9]. Model ini dilatih menggunakan korpus besar yang terdiri dari sekitar 26 juta *tweet* dengan total 409 juta token, yang dikumpulkan melalui API resmi X menggunakan 60 kata kunci berbagai domain seperti ekonomi, kesehatan, pendidikan, dan pemerintahan. Dengan pendekatan berbasis *contextual embedding*, IndoBERTweet mampu merepresentasikan makna kata secara kontekstual, sehingga setiap kata memiliki bobot representasi yang berbeda, tergantung pada konteks kalimatnya. Hal ini menjadikan IndoBERTweet efektif dalam meningkatkan kualitas fitur pada proses klasifikasi sentimen.

2.5. Pembagian Data

Pembagian data dilakukan dengan membagi dataset menjadi 80% data latih dan 20% data uji. Pemilihan rasio ini mengacu pada penelitian sebelumnya yang menujukkan bahwa proporsi 80 : 20 memberikan performa terbaik dalam klasifikasi menggunakan SVM dibandingkan rasio lainnya, sehingga dinilai optimal dalam mengukur kemampuan generalisasi model terhadap data baru [14].

2.6. Klasifikasi Data

Klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) yang merupakan algoritma *supervised learning* yang bekerja dengan memanfaatkan konsep vektor dalam proses klasifikasi. Algoritma ini membangun sebuah hyperplane sebagai garis pemisah antar kelas dalam ruang fitur. *Hyperplane* tersebut ditentukan dengan memaksimalkan jarak anatar kelas, yang dikenal sebagai margin, sehingga menghasilkan pemisah yang optimal [15]. SVM juga dikenal mampu menangani data yang kompleks serta memberikan kinerja yang baik pada dataset berdimensi tinggi [16].

Secara matematis, *hyperplane* dalam SVM didefinisikan sebagai berikut [17]:

$$h(x) = \mathbf{w} \cdot \mathbf{x} + b$$

dimana w adalah vektor yang tegak lurus terhadap *hyperplane*, x merupakan data, dan b adalah bias. Parameter w dan b dihitung

berdasarkan kombinasi linier dari data pelatihan [18].

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$b = -\frac{1}{2}\left(w \cdot x^{+} + w \cdot x^{-}\right)$$

Margin didefinisikan sebagai jarak antara dua *hyperplane* yang membatasi kelas positif dan negatif, yang dirumuskan sebagai berikut [17]:

$$\textit{Margin} = \frac{2}{\|w\|}$$

Proses klasifikasi pada SVM bertujuan untuk menemukan sebuah garis (*hyperplane*) yang mampu memisahkan kedua kelas tersebut. Berbagai alternatif garis pemisah dapat terbentuk, namun SVM memilih *hyperplane* yang memiliki margin paling optimal, yaitu jarak maksimum antara *hyperplane* dengan titik-titik terdekat dari masing-masing kelas yang disebut sebagai *support vectors*. Pada ilustrasi tersebut, *hyperplane* optimal terletak di tengah kedua kelas dengan margin yang maksimal. Pendekatan ini menjadi inti dari mekanisme kerja SVM karena mampu meningkatkan akurasi klasifikasi serta kemampuan generalisasi model terhadap data baru.

Pada kondisi tertentu, data tidak dapat dipisahkan secara linear. Untuk mengatasi hal tersebut, SVM menggunakan teknik *kernel trick*, yaitu mentransformasi data ke ruang fitur berdimensi lebih tinggi sehingga memungkinkan pemisahan linear dilakukan pada ruang tersebut [19]. Tahapan dimulai dari ruang input awal, dimana data dari dua kelas saling bercampur sehingga tidak dapat dipisahkan menggunakan garis linear. Namun setelah dilakukan pemetaan ke ruang berdimensi lebih tinggi, kedua kelas menjadi lebih terstruktur sehingga sebuah hyperplane dapat dibentuk untuk memisahkan keduanya dengan lebih efektif. Proses ini menjadi dasar penggunaan fungsi kernel dalam SVM karena memungkinkan model bekerja pada ruang berdimensi tinggi tanpa perlu menghitung transformasi secara eksplisit [20]. Beberapa jenis kernel yang umum digunakan dalam SVM antara lain [8]:

1. *Linear Kernel*
$$K(x_i, x_j) = x_i \cdot x_j$$
Digunakan untuk data yang dapat dipisahkan secara linier.
2. *Polynomial Kernel*
$$K(x_i, x_j) = (x_i \cdot x_j + 1)^p$$
Digunakan untuk menangani hubungan *non-linear* dengan derajat tertentu.
3. *RBF Kernel*
$$K(x_i, x_j) = \exp\left(-\gamma \, \|x_i - x_j\|^2\right)$$
Kernel ini mampu menangkap pola *non-linear* yang kompleks, namun sensitif terhadap parameter.
4. *Sigmoid Kernel*
$$K(x_i, x_j) = \tanh\left(\gamma \, x_i \cdot x_j + v\right)$$
Kernel yang menggunakan fungsi aktivasi *sigmoid* untuk memodelkan hubungan *non-linear*.

Dalam penelitian ini, digunakan kernel RBF karena memiliki fleksibilitas tinggi dalam menangani variasi distribusi data serta mampu menghasilkan batas keputusan yang optimal pada kasus klasifikasi sentimen. Selain itu, kernel RBF terbukti memberikan kinerja terbaik dibandingkan kernel lainnya dengan tingkat akurasi mencapai 84% [21].

2.7. Evaluasi Model Klasifikasi

Evaluasi merupakan tahap akhir dalam proses klasifikasi yang bertujuan untuk mengukur kinerja model yang telah dibangun [8]. Pada penelitian ini, digunakan algoritma SVM dengan dua label sentimen, yaitu positif dan negatif, sedangkan sentimen netral tidak disertakan karena tidak memberikan kontribusi signifikan dalam pengambilan keputusan klasifikasi.

Untuk menilai performa model, digunakan *confusion matrix* sebagai alat evaluasi yang mampu memberikan gambaran menyeluruh mengenai kualitas prediksi model terhadap data uji yang telah diketahui label sebenarnya. Melalui *confusion matrix*, berbagai metrik evaluasi seperti *accuracy, precision, recall*, dan *f1-score* dapat dihitung sehingga analisis kinerja model menjadi lebih komprehensif [22]. Tabel 1

Table 1. *Confusion Matrix*

Actual	*Prediction*	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Berdasarkan hasil *confusion matrix* diperoleh empat metrik evaluasi sebagai berikut:

1. *Accuracy*
Hitung rasio prediksi yang benar terhadap jumlah total titik data uji.
$$\textit{Accuracy} = \frac{(TP + TN + TNt)}{(TP + TN + TNt + FP + FN + FNt)} \times 100\%$$
2. *Precision*
Menunjukkan ketepatan model dalam memprediksi kelas positif.
$$\textit{Precision} = \frac{TP}{TP + FP}$$
3. *Recall*
Menilai kemampuan model dalam mengidentifikasi semua data positif secara akurat.
$$\textit{Recall} = \frac{TP}{TP + FN}$$
4. *F1-Score*
Menunjukkan keseimbangan antara *precision* dan *recall*.
$$\textit{F1-Score} = 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

3. Hasil dan Pembahasan

3.1. Preprocessing data

Penelitian ini menggunakan data berupa komentar dan tanggapan pengguna media sosial X yang dihimpun selama periode 10 Oktober hingga 10 Januari 2025. Dengan menggunakan

kata kunci "redenominasi rupiah", sebanyak 1.014 tweet berhasil ditarik menggunakan *package tweet-harvest* pada bahasa pemrograman *Python*. Seluruh data tersebut kemudian disimpan dalam format *Microsoft Excel*. Contoh hasil pengumpulan data dapat dilihat pada tabel 2

Table 2. Contoh Hasil Pengumpulan Data

No.	*Tweet*
1	@paranoia365852 Akan ada sesuatu besar dengan 6 mata uang ini terhadap USD. Itulah mengapa Rupiah di redenominasi. Semuanya ini tentang currency reset.
2	@yatever @vespamatic96 Rencana redenominasi rupiah memang ada dalam Renstra Kemenkeu 2025–2029, tapi hingga Januari 2026 masih dalam tahap kajian dan belum dilaksanakan. Target penyelesaian RUU pada 2027 tergantung stabilitas ekonomi.
3	@pratagumala @prabowo Ini sih tinggal redenominasi aja rupiahnya jadi kurang satu nolnya. Jadi deh 1 USD = 1.670 rupiah sekarang.

Data cuitan diproses melalui tahap *pre-processing* agar siap diklasifikasikan. Cuitan duplikat dihapus untuk menjaga kualitas data, sehingga diperoleh 376 cuitan untuk analisis. Tahapan *preprocessing* meliputi:

a. *Case folding*: mengubah seluruh huruf menjadi huruf kecil.
b. *Cleaning*: menghapus karakter khusus dan elemen teks yang tidak relevan.
c. *Tokenization*: memecah teks menjadi kata-kata (*token*).
d. *Normalization*: mengubah kata tidak baku menjadi bentuk baku.
e. *Stemming*: mengubah kata menjadi bentuk dasarnya.

Hasil dari preprocessing dapat dilihat pada tabel 3

3.2. Pelabelan Data

Tahap selanjutnya adalah klasifikasi sentimen melalui pelabelan manual pada setiap cuitan oleh penulis. Sentimen dikelompokkan menjadi tiga kategori, yaitu positif (1), netral (0), dan negatif (-1). Label positif diberikan pada cuitan yang menunjukkan dukungan, sedangkan label negatif diberikan pada cuitan yang mengandung penolakan. Cuitan yang tidak menunjukkan kecenderungan positif atau negatif dikategorikan sebagai netral.

Hasil pelabelan kemudian divalidasi oleh ahli bahasa Indonesia. Berdasarkan validasi, diperoleh 226 cuitan positif, 644 cuitan netral, dan 144 cuitan negatif. Penelitian ini hanya menggunakan dua kelas sentimen, yaitu positif dan negatif, sedangkan data netral dieliminasi karena tidak relevan dengan tujuan pengambilan keputusan. Dengan demikian, data yang digunakan pada tahap analisis selanjutnya berjumlah 370 cuitan, terdiri atas 226 sentimen positif dan 144 sentimen negatif. Hasil dari pelabelan data yang telah divalidasi dapat dilihat pada tabel 4:

3.3. IndoBERTweet

Tahapan selanjutnya adalah ekstraksi fitur untuk mengubah teks menjadi representasi numerik agar dapat diproses oleh algoritma *machine learning*. Pada penelitian ini menggunakan model *pre-trained* IndoBERTweet, yaitu model berbasis *transformer* yang dilatih pada data media sosial berbahasa Indonesia, sehingga mampu menangkap kata secara lebih akurat dalam bentuk vektor (*embedding*) berdimensi 768. Berdasarkan tabel **??**, dimensi vektor fitur yang dihasilkan berjumlah 768 sesuai dengan konfigurasi standar (*default*) dari model IndoBERTweet. Vektor hasil ekstraksi tersebut kemudian akan melalui tahapan seleksi fitur. Proses ini bertujuan untuk mereduksi dimensi yang dianggap tidak relevan guna meningkatkan efisiensi proses klasifikasi.

3.4. Pembagian Data

Setelah fitur diekstraksi, data dengan label netral dieliminasi. Hal ini dilakukan karena data netral sebelumnya hanya berisi informasi yang tidak menyatakan dukungan ataupun penolakan sehingga tidak berpengaruh dalam konteks pengambilan keputusan. Sehingga jumlah data yang digunakan sebanyak 370 yang terdiri dari 226 data positif dan 144 data negatif. Selanjutnya, data tersebut dibagi menjadi dua bagian, yaitu data latih dan data uji, dengan rasio perbandingan 80:20. Sebanyak 80% data (296 *tweet*) digunakan sebagai data latih, sedangkan 20% data (74 *tweet*) digunakan sebagai data uji.

3.5. Klasifikasi SVM

Hasil klasifikasi sentimen menggunakan SVM dengan parameter $C = 10$ dan $\gamma = scale$ disajikan dalam bentuk tabel 6:

Table 6. *Confusion Matrix* SVM dengan Ekstraksi Fitur IndoBERTweet

Label	Positif	Negatif
Positif	23	6
Negatif	7	38

Berdasarkan *confusion matrix*, maka dihasilkan hasil *accuracy, precision, recall,* dan *f1-Score* disajikan pada tabel 7.

Table 7. Hasil Evaluasi Model SVM

Kelas	*Precision*	*Recall*	*F1-Score*	*Accuracy*
Negative	0,77	0,79	0,78	
Positive	0,86	0,84	0,85	82%
Average	0,82	0,82	0,82	

Hasil pengujian menunjukkan bahwa model SVM memperoleh nilai akurasi sebesar 82%, yang mengindikasikan bahwa model mampu mengklasifikasikan sentimen positif dan negatif dengan cukup baik, di mana sebagian besar data sesuai dengan label aktual. Nilai *precision* rata-rata sebesar 82% menunjukkan bahwa sebagian besar hasil prediksi sudah tepat, sedangkan *recall* sebesar 82% menunjukkan kemampuan model dalam mengenali sebagian besar data yang relevan. Sementara itu, *f1-score* sebesar 82% mencerminkan keseimbangan yang baik antara *precision* dan *recall*, sehingga model cukup konsisten dalam melakukan klasifikasi.

Namun, performa model belum merata pada setiap kelas, khususnya pada kelas negatif. Hal ini ditunjukkan oleh masih adanya kesalahan klasifikasi berupa *false negative* dan *false positive*, yang mengindikasikan bahwa model belum optimal dalam mengenali seluruh pola sentimen negatif. Kondisi tersebut kemungkinan disebabkan oleh karakteristik data media

Table 3. Hasil Tahapan *Preprocessing* data

Proses	Input	Output
Case Folding	@paranoia365852 Akan ada sesuatu besar dengan 6 mata uang ini terhadap USD. Itulah mengapa Rupiah di redenominasi. Semuanya ini tentang currency reset	@paranoia365852 akan ada sesuatu besar dengan 6 mata uang ini terhadap usd. itulah mengapa rupiah di redenominasi. semuanya ini tentang currency reset
Cleaning	@paranoia365852 akan ada sesuatu besar dengan 6 mata uang ini terhadap usd. itulah mengapa rupiah di redenominasi. semuanya ini tentang currency reset	akan ada sesuatu besar dengan mata uang ini terhadap usd itulah mengapa rupiah di redenominasi semuanya ini tentang currency reset
Tokenizing	akan ada sesuatu besar dengan mata uang ini terhadap usd itulah mengapa rupiah di redenominasi semuanya ini tentang currency reset	['akan', 'ada', 'sesuatu', 'besar', 'dengan', 'matauang', 'ini', 'terhadap', 'usd', 'itulah', 'mengapa', 'rupiah', 'di', 'redenominasi', 'semuanya', 'ini', 'tentang', 'currency', 'reset']
Normalization	['akan', 'ada', 'sesuatu', 'besar', 'dengan', 'matauang', 'ini', 'terhadap', 'u', 'd', 'itulah', 'mengapa', 'rupiah', 'di', 'redenominasi', 'semuanya', 'ini', 'tentang', 'currency', 'reset']	['akan', 'ada', 'sesuatu', 'besar', 'dengan', 'matauang', 'ini', 'terhadap', 'kamu', 'di', 'itulah', 'mengapa', 'rupiah', 'di', 'redenominasi', 'semuanya', 'ini', 'tentang', 'currency', 'reset']
Stemming	['akan', 'ada', 'sesuatu', 'besar', 'dengan', 'matauang', 'ini', 'terhadap', 'kamu', 'di', 'itulah', 'mengapa', 'rupiah', 'di', 'redenominasi', 'semuanya', 'ini', 'tentang', 'currency', 'reset']	akan ada sesuatu besar dengan matauang ini hadap kamu di itulah mengapa rupiah di redenominasi semua ini tentang currency reset

Table 4. Contoh Hasil Pelabelan dan Validasi Data

No.	*Tweet*	Label Awal	Label Validasi	Catatan
1	akan ada sesuatu besar dengan matauang ini hadap usd di itu mengapa rupiah di redenominasi semua ini tentang currency reset.	0	1	Berubah karena tweet mengandung dukungan berupa promosi redenominasi rupiah
2	rencana redenominasi rupiah memang ada dalam rencana strategis kemenkeu tapi hingga januari masih dalam tahap kaji dan belum laksana target selesai ruu pada gantung stabilitas ekonomi	0	0	-
3	ini sih tinggal redenominasi saja rupiah jadi kurang satu nol jadi deh usd rupiah sekarang	0	0	-

sosial yang tidak terstruktur, seperti penggunaan bahasa informal, singkatan, sarkasme, serta ambiguitas konteks. Selain itu, distribusi data yang tidak seimbang juga memengaruhi kemampuan model dalam mengenali sentimen negatif secara optimal.

4. Kesimpulan

Penelitian ini dilakukan untuk menganalisis dan mengklasifikasikan sentimen opini publik terhadap isu redenominasi rupiah di media sosial X dengan memanfaatkan pendekatan machine learning. Permasalahan utama yang dihadapi adalah karakteristik data teks media sosial yang tidak terstruktur, penggunaan bahasa informal, serta adanya ambiguitas konteks yang dapat memengaruhi hasil klasifikasi. Berdasarkan hasil penelitian, penggunaan metode SVM dengan ekstraksi fitur IndoBERTweet mampu menghasilkan performa klasifikasi yang cukup baik. Model yang dibangun memperoleh nilai akurasi sebesar 82%, *precision* 82%, *recall* 82%, dan *f1-score* 82%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan sentimen positif dan negatif secara seimbang dan cukup konsisten. Selain itu, penggunaan IndoBERTweet terbukti efektif dalam menghasilkan representasi vektor berdimensi 768 yang mampu menangkap makna kontekstual bahasa Indonesia, termasuk bahasa informal yang umum digunakan di media sosial, sehingga mendukung peningkatan kinerja klasifikasi SVM.

Hasil klasifikasi juga menunjukkan bahwa 61% opini publik bersifat positif dan 39% bersifat negatif. Hal ini mengindikasikan bahwa mayoritas masyarakat cenderung memberikan respons positif terhadap isu redenominasi rupiah, meskipun masih terdapat sebagian opini negatif yang berkaitan dengan kekhawatiran terhadap dampak ekonomi. Adapun saran untuk penelitian selanjutnya adalah memperluas jumlah dan variasi data agar model dapat belajar dari distribusi data yang lebih beragam. Selain itu, penelitian berikutnya dapat mempertimbangkan penggunaan metode *deep learning* lain lainnya untuk meningkatkan performa klasifikasi.

References

[1] M. A. U. Hasan, A. A. Bakar, and M. R. Yaakub, "Generating attribute similarity graphs: A user behavior-based approach from real-time microblogging data on platform x," *Research Square*, 2024.

[2] M. A. Faidh, M. E. Maulana, N. E. Putri, S. I. Putri, T. A. Munir, and A. Laksana, "Peran media sosial x dalam perkembangan komunikasi di era digital," *Konsensus: Jurnal Ilmu Pertahanan, Hukum dan Ilmu Komunikasi*, vol. 1, no. 6, pp. 43–51, 2024.

[3] T. N. Wijaya, R. Indriati, and M. N. Muzaki, "Analisis sentimen opini publik tentang undang-undang cipta kerja pada twitter," *Jambura Journal of Electrical and Electronics Engineering*, vol. 3, no. 2, pp. 78–83, 2021.

[4] H. C. Husada and A. S. Paramita, "Analisis sentimen pada maskapai penerbangan di platform twitter menggunakan algoritma support vector machine (svm)," *Teknika*, vol. 10, no. 1, pp. 18–26, 2021.

[5] A. D. A. Putra and S. Juanita, "Analisis sentimen pada ulasan pengguna aplikasi bibit dan bareksa dengan algoritma knn," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, pp. 636–646, 2021.

[6] A. S. Aribowo, H. Basiron, N. F. Abd Yusof, and S. Khomsah, "Cross-domain sentiment analysis model on indonesian youtube comment." *International Journal of Advances in Intelligent Informatics*, vol. 7, no. 1, 2021.

[7] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis sentimen terhadap penggunaan aplikasi shopee mengunakan algoritma support vector machine (svm)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, 2023.

[8] T. M. P. Aulia, N. Arifin, and R. Mayasari, "Perbandingan kernel support vector machine (svm) dalam penerapan analisis sentimen vaksinisasi covid-19," *SINTECH (Science and Information Technology) Journal*, vol. 4, no. 2, pp. 139–145, 2021.

[9] F. Koto, J. H. Lau, and T. Baldwin, "Indobertweet: A pretrained language model for indonesian twitter with effective domain-specific vocabulary initialization," *arXiv preprint arXiv:2109.04607*, 2021.

[10] M. D. Rahman, A. Djunaidy, and F. Mahananto, "Penerapan weighted word embedding pada pengklasifikasian teks berbasis recurrent neural network untuk layanan pengaduan perusahaan transportasi," *Jurnal Sains dan Seni ITS*, vol. 10, no. 1, pp. A1–A6, 2021.

[11] J. W. Iskandar, Y. Nataliani *et al.*, "Perbandingan naïve bayes, svm, dan k-nn untuk analisis sentimen gadget berbasis aspek," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021.

[12] H. T. A. Simanjuntak, P. E. P. Silaban, J. K. S. Manurung, and V. H. Sormin, "Klasterisasi berita bahasa indonesia dengan menggunakan k-means dan word embedding," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 3, pp. 641–652, 2023.

[13] L. Hermawan and M. B. Ismiati, "Pembelajaran text preprocessing berbasis simulator untuk mata kuliah information retrieval," *Jurnal Transformatika*, vol. 17, no. 2, pp. 188–199, 2020.

[14] M. Muadin, J. Junadhi, R. Rahmiati, and H. Asnal, "Implementasi metode support vector machine pada opinion mining masyarakat terkait chatgpt," *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 7, no. 1, pp. 78–84, 2023.

[15] Z. Arifin, D. F. Rahman, B. S. Rintyarna, and D. Daryanto, "Penerapan algoritma support vector machine berbasis kernel radial basis function dalam klasifikasi sel kanker," *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 4, no. 2, pp. 100–106, 2023.

[16] H. Eldo, A. Ayuliana, D. Suryadi, G. Chrisnawati, and L. Judijanto, "Penggunaan algoritma support vector machine (svm) untuk deteksi penipuan pada transaksi online," *Jurnal Minfo Polgan*, vol. 13, no. 2, pp. 1627–1632, 2024.

[17] F. Alvianda, I. Indriati, and P. P. Adikara, "Analisis sentimen konten radikal di media sosial twitter menggunakan metode support vector machine (svm)," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 3, no. 1, pp. 241–246, 2019.

[18] A. Susilo, "Analisis sentimen sara pada tweet berbahasa indonesia menggunakan indobert dan support vector machine (svm)," digilib.uinsa.ac.id, 2021, skripsi, Fakultas Sains dan Teknologi, Jurusan Teknik Informatika.

[19] M. A. A. Syukur, M. H. Susanto, and S. Alfarizhi, "Klasifikasi kualitas air dengan menggunakan metode support vector machine," *Maliki Interdisciplinary Journal*, vol. 2, no. 11, pp. 1288–1299, 2024.

[20] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran kinerja sistem kualitas udara dengan teknologi wsn menggunakan confusion matrix," *Jurnal Informatika Upgris*, vol. 6, no. 2, 2020.

[21] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan evaluasi kernel svm untuk klasifikasi sentimen dalam analisis kenaikan harga bbm: Comparative evaluation of svm kernels for sentiment classification in fuel price increase analysis," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, 2023.

[22] S. Lonang, A. Yudhana, and M. K. Biddinika, "Analisis komparatif kinerja algoritma machine learning untuk deteksi stunting," *J. Media Inform. Budidarma*, vol. 7, no. 4, p. 2109, 2023.