

## Integration of BERT and SVM in Sentiment Analysis of Twitter/X Regarding Constitutional Court Decision No. 60/PUU-XXII/2024

Artia Irianti<sup>\*1</sup>, Halimah<sup>2</sup>, Sutedi<sup>3</sup>, Melda Agariana<sup>4</sup>

<sup>1,2,3,4</sup>Department of Information System, IIB Darmajaya, Indonesia

Email: [iriantiarial1@gmail.com](mailto:iriantiarial1@gmail.com)

Received : Oct 31, 2024; Revised : Dec 15, 2024; Accepted : Dec 30, 2024; Published : Apr 26, 2025

### Abstract

This research analyzes public sentiment towards the Indonesian Constitutional Court's decision No. 60/PUU-XXII/2024 by utilizing natural language processing techniques using the BERT (Bidirectional Encoder Representations from Transformers) model and the Support Vector Machine model (SVM). The research methodology includes four stages: data preprocessing, data labeling using BERT, embedding extraction, and SVM model training. The data is taken from the Twitter platform, where various public opinions are reflected in three sentiment categories: positive, neutral, and negative. The preprocessing process results in the removal of approximately 23% of duplicate data, and sentiment labeling shows a dominance of the positive category. Evaluation results from the SVM model training demonstrated varying performance: negative sentiment achieved a Precision of 0.57, Recall of 0.36, and F1-score of 0.44; neutral sentiment had a Precision of 0.81, Recall of 0.62, and F1-score of 0.70; while positive sentiment recorded a Precision of 0.98, Recall of 1.00, and F1-score of 0.99. The model's overall accuracy reached 0.97. These findings indicate that the integration of BERT and SVM is effective for sentiment classification, but improvements are needed in the negative and neutral categories to achieve more balanced performance.

**Keywords :** *Machine Learning , Natural Language Processing (NLP), Public Opinion, Social Media, Text Classification.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



## 1. INTRODUCTION

Berbicara tentang hukum, salah satu produk hukum di Indonesia adalah putusan Mahkamah Konstitusi atau yang biasa kita sebut MK. Jika kita membahasa tentang MK, lembaga ini baru saja mengeluarkan produk hukum berupa putusan No.60/PUU-XXII/2024[1]. Dari putusan tersebut berdampak pada opini publik yang mana melahirkan pendapat, kritik, dan dukungan. Dampak dari putusan tersebut menarik perhatian masyarakat Indonesia yang menimbulkan reaksi dari berbagai kalangan. Bentuk dari reaksi masyarakat tersebut dapat dilihat dengan meningkatnya aktivitas media sosial yang berhubungan dengan putusan MK tersebut[2]. Putusan yang menarik perhatian masyarakat dan memicu beragam reaksi di media sosial tersebut dapat dianalisis lebih lanjut melalui proses sentimen analisis, yang bertujuan untuk mengenali dan mengklasifikasikan pendapat-pendapat yang diungkapkan dalam teks[3], [4]. Sentimen analisis adalah proses untuk mengenali dan mengklasifikasikan pendapat yang disampaikan dalam teks, terutama di media sosial, atau untuk tujuan klasifikasi[5]. Adapun tujuan lain yang ingin dicapai pada penelitian ini adalah ingin memastikan apakah proses pelatihan sebuah model dapat dilakukan dengan bantuan proses *embedding* pada model lain, yang artinya dalam proses tersebut terdapat dua model yang akan digunakan.

Sentimen analisis juga merupakan proses ekstraksi opini, emosi individu, sikap, dan pandangan dari teks atau ucapan yang terdapat dalam konten web seperti *tweet* menggunakan *Natural Language Processing* (NLP)[6]. Selain NLP sentimen dapat dilakukan dengan cara menerapkan model *machine*

*learning* pada sebuah data teks. *Machine learning* (pembelajaran mesin) didefinisikan sebagai jenis kecerdasan buatan (AI)[7], [8] yang memungkinkan mesin untuk dilatih tanpa harus diprogram secara eksplisit[9].

BERT merupakan salah satu model yang dapat digunakan dalam melakukan analisis sentimen, BERT sendiri merupakan singkatan dari *Bidirectional Encoder Representations from Transformers*, yang merupakan model representasi bahasa yang dirancang untuk melakukan *pre-training* representasi *bidirectional* yang dalam, dari teks yang tidak memiliki label dengan mengkondisikan secara bersamaan pada konteks kiri dan kanan di semua lapisan. Hasilnya, model BERT yang telah di *pre-train* dapat di *fine-tune* dengan hanya menambahkan satu lapisan *output* tambahan untuk menciptakan model-model yang mencapai hasil terbaik pada berbagai tugas, seperti *question answering* dan *language inference*, tanpa memerlukan modifikasi arsitektur yang spesifik untuk tugas tersebut. BERT memiliki arsitektur yang seragam untuk berbagai tugas dan menggunakan mekanisme *self-attention* untuk memodelkan banyak tugas *downstream*, baik yang melibatkan teks tunggal maupun pasangan teks[10] model berbasis BERT juga terbukti memberikan hasil terbaik dengan nilai akurasi yang lebih tinggi daripada model lainnya[11]. Integrasi BERT telah terbukti efektif dalam berbagai tugas pemrosesan bahasa alami, termasuk klasifikasi teks, [12] Model BERT yang dikembangkan menggunakan SVM juga menghasilkan akurasi cukup tinggi [13]. Model SVM (*Support Vector Machine*) adalah model klasifikasi biner yang berfungsi sebagai pengklasifikasi *linear*. Model ini bekerja dengan menemukan *hyper-plane* segmentasi optimal yang tidak hanya memisahkan dua kelas data dengan benar, tetapi juga memaksimalkan margin klasifikasi antara kedua kelas data tersebut [14], [15].

Pada penelitian ini BERT dipilih karena kemampuannya dalam memahami konteks teks secara mendalam melalui representasi *bidirectional*. Setelah data diberi label berdasarkan keluaran model BERT, data berlabel tersebut digunakan sebagai data pelatihan untuk model SVM (*Support Vector Machine*). Model SVM diterapkan sebagai pengklasifikasi *linear* untuk membedakan kategori sentimen dalam data. Penggunaan SVM setelah pelabelan dengan BERT diharapkan dapat meningkatkan akurasi klasifikasi dengan memanfaatkan kekuatan BERT dalam pemrosesan bahasa alami dan efisiensi SVM dalam pengklasifikasian data. Integrasi BERT dan SVM dalam analisis sentimen Twitter diharapkan mampu menawarkan pendekatan yang lebih akurat untuk menangkap nuansa sentimen masyarakat terhadap putusan MK. Dengan menggabungkan kemampuan BERT dalam memahami konteks bahasa yang kompleks dengan kekuatan SVM dalam mengklasifikasikan data, diharapkan menghasilkan model klasifikasi sentimen yang lebih akurat, mampu memahami nuansa bahasa secara lebih baik.

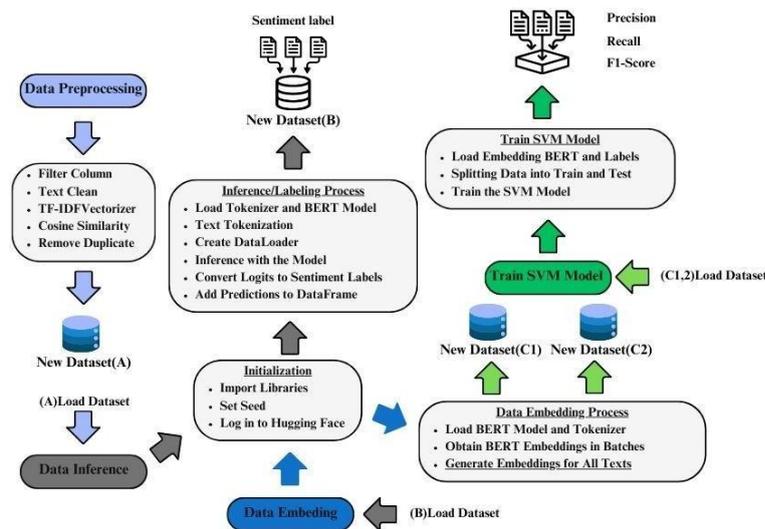
Adapun penelitian sebelumnya model BERT digunakan dalam analisis sentimen, dan pada penelitian tersebut dilakukan pendekatan tradisional menggunakan model BERT berbasis bahasa Arab model QARIB hasil dari penelitian tersebut model QARIB mencapai akurasi tinggi sebesar 96% bahkan dengan data yang tidak seimbang[16]. Model IndoBERT sendiri memiliki kinerja yang lebih baik dibandingkan dengan metode klasik seperti SVM dan *Multinomial Naive Bayes*[17]. Selain itu BERT juga dapat digunakan sebagai model dasar untuk deteksi ujaran kebencian dan mitigasi bias rasial di media sosial[18]. BERT juga pernah digunakan untuk klasifikasi sentimen pada postingan Weibo terkait COVID-19 di China[19]. BERT juga digunakan untuk analisis sentimen dan deteksi entitas kunci pada teks keuangan *online*[20].

Model BERT juga telah digunakan untuk klasifikasi opini publik mengenai *coronavirus*, dan hasil eksperimen menunjukkan bahwa akurasi validasi dari klasifikasi emosi mencapai sekitar 94%[21]. Penerapan BERT digunakan sebagai model *embedding* kata yang berfungsi sebagai *encoder*, hal ini digunakan untuk mendapatkan representasi konteks dari sebuah kalimat secara akurat yang kemudian diterapkan dalam kombinasi dengan blok paralel dari *Convolutional Neural Network* (CNN)

satu lapis untuk mendeteksi berita palsu[22]. SVM sendiri merupakan algoritma yang kerap digunakan sebagai metode pembelajaran mesin yang di supervisi/diawasi untuk mengklasifikasikan data berdasarkan sentimen positif, negatif, dan netral[23]. *Support Vector Machine* (SVM) juga bisa digunakan untuk analisis sentimen teks dengan memanfaatkan fungsi kernel *fisher* yang didasarkan pada model *Probabilistic Latent Semantic Analysis* (PLSA). Metode ini memungkinkan informasi semantik laten yang melibatkan karakteristik probabilitas digunakan sebagai fitur klasifikasi, sehingga meningkatkan efektivitas klasifikasi SVM. Metode ini juga mengatasi masalah pengabaian karakteristik semantik laten dalam analisis sentimen teks[24]. *Support Vector Machine* (SVM) juga digunakan sebagai bagian dari *ensemblar evolusioner* yang dirancang untuk melakukan klasifikasi multi-label pada teks. SVM diintegrasikan dengan algoritma genetik (GA) untuk meningkatkan akurasi/ ketepatan, presisi/ kecermatan, *recall*/daya ingat, dan *f-measure*/ukuran harmonis dari klasifikasi multi-label berbasis aspek[25].

## 2. METHOD

Penelitian ini melalui empat tahapan proses yaitu: *preprosesing* data, pelabelan data, *embedding* data dan yang terakhir adalah pelatihan model SVM dari hasil *embedding* data. Terlihat pada Gambar 1



Gambar 1. Empat tahapan proses pada penelitian

### 2.1. Preprosesing Data

Tahapan ini bertujuan membersihkan teks, menghitung kemiripan antar teks menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) dan *cosine similarity*, serta menghapus teks duplikat. Setelah dibersihkan, data disimpan ke *dataset* baru untuk pelabelan dengan model BERT. Langkah-langkahnya meliputi: memuat dan memfilter *dataset* dengan memilih kolom seperti *created\_at*, *full\_text*, statistik interaksi, dan *username*; membersihkan teks dengan menghilangkan elemen seperti URL dan tanda baca; menerapkan TF-IDF untuk mengukur bobot kata; menghitung kemiripan antar teks dengan *cosine similarity* (*threshold* 0.8); serta menghapus teks duplikat, mempertahankan teks pertama.

### 2.2. Pelabelan Data Dengan BERT/Inferensi Data

Model IndoBERT yang digunakan adalah model indobenchmark/indobert-base-p1 sedangkan parameter yang digunakan pada model tersebut adalah, menggunakan 12 *attention heads* pada 12 lapisan *encoder (hidden layers)* untuk memproses informasi, dengan jumlah representasi (*hidden size*) sebanyak 768 lalu panjang token maksimum: 512 token dan menggunakan fungsi aktivasi *Gaussian Error Linear Unit* (GELU) dengan probabilitas *dropout* sebesar 0,1 untuk mencegah *overfitting*, baik pada *hidden states* maupun *attention probabilities*. Parameter lain seperti *pooler\_type* dilakukan berdasarkan token pertama selanjutnya model tidak diatur untuk tugas *generation* karena parameter *is\_encoder\_decoder* bernilai *false*, *vocabulary size* sebanyak 50.000 kata format data yang digunakan adalah *float32* pada parameter *decoding* tidak ada penalti untuk pengulangan token selama *decoding repetition\_penalty*: 1.0[26]

Proses inferensi sentimen dengan BERT diawali dengan mengatur *seed* untuk hasil yang konsisten dan memuat data teks bersih ke dalam *DataFrame*. Model dan *tokenizer* diinisialisasi dari model pra-latih IndoBERT dengan tiga label sentimen: Negatif, Netral, dan Positif, dan dijalankan di GPU (Graphics Processing Unit) jika tersedia. Teks kemudian ditokenisasi dengan memperhatikan panjang maksimum dan token khusus, lalu dibagi menjadi *batch* menggunakan *DataLoader*. Setiap *batch* diproses oleh model untuk menghasilkan *logits*, yang kemudian diubah menjadi prediksi label sentimen berdasarkan nilai tertinggi ukuran *batch* pada proses ini sebesar 128 sehingga 128 sampel akan diproses bersama dalam setiap iterasi. Hasil akhir ditambahkan ke *DataFrame* dan disimpan sebagai file CSV baru.

### 2.3. Embedding Data

Proses ekstraksi *embeddings* dari BERT diawali dengan *login* ke *Hugging Face Hub* menggunakan token untuk mengakses model IndoBERT. Kemudian, data teks berlabel sentimen dimuat dari file CSV menggunakan *pandas* sebagai *input*. *Tokenizer* dan model IndoBERT diinisialisasi, di mana *tokenizer* mengonversi teks menjadi token yang dapat diproses, dan model BERT menghasilkan representasi numerik atau *embeddings*. Setiap *batch* teks ditokenisasi dan diproses oleh model untuk memperoleh *output* dari lapisan *pooler*, menghasilkan vektor tetap yang mewakili makna teks. *Embeddings* disimpan dalam matriks fitur x, sementara label sentimen disimpan dalam vektor y. Matriks *embeddings* kemudian disimpan dalam file *numpy (.npy)*, sedangkan label disimpan dalam file CSV.

Adapun parameter yang digunakan pada proses ini adalah ukuran *batch* sebesar 32 dan Panjang maksimum setiap teks setelah diproses 128 selebihnya akan di potong.

### 2.4. Pelatihan Model SVM

Tahap penggunaan *embeddings* sebagai fitur untuk SVM dimulai dengan memuat *library* yang diperlukan, seperti *Numpy*, *Pandas*, dan *Scikit-learn*, untuk pemrosesan data dan klasifikasi. *Embedding* BERT yang telah diproses dan label sentimen kemudian diambil dari file yang disimpan, di mana *embedding* berfungsi sebagai representasi numerik teks.

Data dibagi menjadi dua bagian, yaitu *training* set (90% dari data) dan *test* set (10% dari data), menggunakan fungsi *train\_test\_split*. Parameter *test\_size=0.1* ini menentukan proporsi data uji, sedangkan *random\_state=42* memastikan pembagian data dilakukan secara konsisten setiap kali siklus kode dijalankan, sehingga hasil eksperimen dapat direproduksi dari jumlah 14645 data.

Model SVM dilatih menggunakan fungsi SVC dengan kernel 'rbf' (*Radial Basis Function*), setelah model dilatih, prediksi dilakukan pada data uji, menghasilkan laporan evaluasi dengan metrik seperti presisi, *recall*, dan *f1-score*. Selanjutnya untuk pertimbangan evaluasi model maka dilakukan proses *k-fold cross-validation* untuk mengevaluasi model SVM yang menggunakan *embedding* BERT sebagai fitur input adapun parameter yang digunakan pada proses *cross-validation* adalah *n\_splits=5*

berarti *dataset* dibagi menjadi 5 , lalu *shuffle=true* memastikan bahwa data diacak sebelum dibagi dan *random\_state=42* untuk memberikan konsistensi pada hasil.

### 3. RESULT

#### 3.1. Hasil

##### 3.1.1. Preprocessing Data

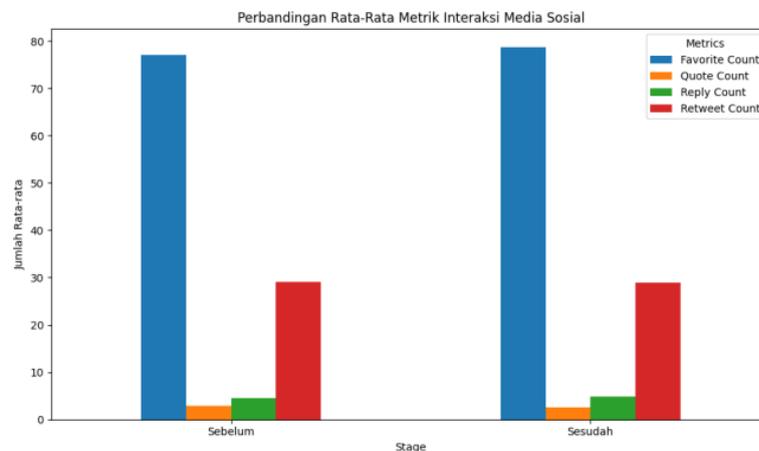
Pada proses pembersihan, beberapa langkah telah dilakukan diantaranya: menghapus URL, *mentions* (@), *hashtags* (#), angka, dan tanda baca, mengubah semua teks menjadi huruf kecil, menghapus spasi berlebih. Langkah selanjutnya diikuti dengan proses penghapusan teks yang mirip, menggunakan TF-IDF *vectorizer* dan *cosine similarity* dengan *similarity* > 0.8/80%.

Dataframe Info:

Column	Non-Null Count	Dtype
created_at	19081	object
full_text	19081	object
favorite_count	19081	int64
quote_count	19081	int64
reply_count	19081	int64
retweet_count	19081	int64
username	19081	object
Memory Usage		6.53 MB

Gambar 2. Dataset awal

Ukuran *dataset* sebelum pembersihan adalah 19.081 pada Gambar 2. baris data pada *dataset* asli, lalu ukuran *dataset* setelah dibersihkan menjadi 14.645 baris data, maka jumlah baris yang dihapus adalah 4.436 baris data. Proses pembersihan data mengidentifikasi baris-baris yang memiliki kemiripan teks yang sangat tinggi (dengan ambang batas kemiripan > 80%), dan baris-baris tersebut dihapus untuk menghindari duplikasi informasi. Akibatnya, sekitar 23% dari data asli dihapus. Pembersihan data pada metrik interaksi media sosial menghasilkan perubahan pada rata-rata statistik. Rata-rata jumlah favorit meningkat dari 77,07 menjadi 78,64, sementara rata-rata jumlah kutipan menurun dari 2,85 menjadi 2,59. Rata-rata balasan naik dari 4,60 menjadi 4,84, dan rata-rata *retweet* mengalami sedikit penurunan dari 29,12 menjadi 28,90.



Gambar 3. Metrik interaksi media sosial

### 3.1.2. Pelabelan Data Dengan BERT/Inferensi Data

Hasil klasifikasi sentimen dari data teks yang dianalisis menunjukkan bahwa sentimen terbagi ke dalam tiga kategori utama. Sebagian besar teks, sebanyak 13.872 diklasifikasikan memiliki sentimen positif, menjadikannya kategori dominan. Sentimen netral terdapat pada 642 teks, yang berada pada posisi tengah dalam jumlah distribusinya.

	full_text	sentiment
6335	kawal jam	Positive
9562	kawalputusan mk	Positive
8608	kawall putusan mk	Positive
3078	kawal putusan mk aja	Positive
8593	aksi kawal putusan mk	Positive
5406	mari kawal putusan mk	Positive
11848	soto kawal putusan mk	Positive
13382	peta kawal putusan mk	Positive
7303	kawal putusan mk wkwk	Positive
2233	kawal putusan mk bang	Positive

Gambar 4. Hasil pelabelan sentimen positif

	full_text	sentiment
1904	kawal putusanmk	Neutral
1665	putusan mk pannnn	Neutral
2422	fokus putusan mk cok	Neutral
11990	besok gt putusan mk dan	Neutral
4504	putusan mk berlaku titik	Neutral
13543	ini gabuka di bandung wil	Neutral
11115	kawal putusan mk samasama	Neutral
10259	kawal putusan mk kami butuh	Neutral
1406	woi putusan mk di anulirrr	Neutral
2392	fokus putusan mk kawankawan	Neutral

Gambar 5. Hasil pelabelan sentimen netral

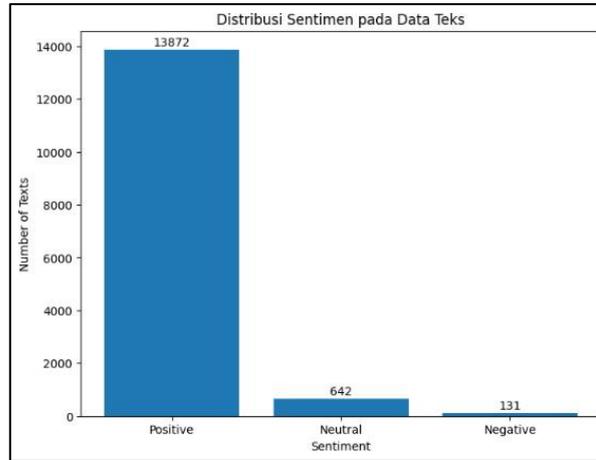
Sementara itu, sentimen negatif muncul pada 131 teks, menjadikannya kategori dengan jumlah paling sedikit.

	full_text	sentiment
8501	aku mencoba kebal lop	Negative
11827	sampai culas ini mematuhi	Negative
12041	pakuwon jogja gaada kah mint	Negative
14383	ini pada masih kecil kecil fak	Negative
3103	putusan mk ngerusak demokrasi brok	Negative
10223	bertentangan dengan putusan mk dong	Negative
11026	berarti tidak ikuti putusan mk simpel	Negative
10445	mbulet menolak revisi putusan mk tidak	Negative
3506	untuk apa ada polling putusan mk mengikat	Negative
5359	semua ln wajib menghormati putusan mk dul	Negative

Gambar 6. Hasil pelabelan sentimen negatif

Dari Gambar 77, dapat diketahui bahwa *dataset* yang digunakan didominasi oleh teks-teks yang bernada positif, dengan teks netral dan negatif berada dalam proporsi yang jauh lebih kecil. *Seed* yang

di atur menjadi 100 berhasil melakukan pelabelan dengan konsisten pada tahapan ini. Model BERT dijalankan dalam mode evaluasi untuk mencegah perubahan pada parameter.



Gambar 7. Distribusi jumlah sentimen yang dihasilkan

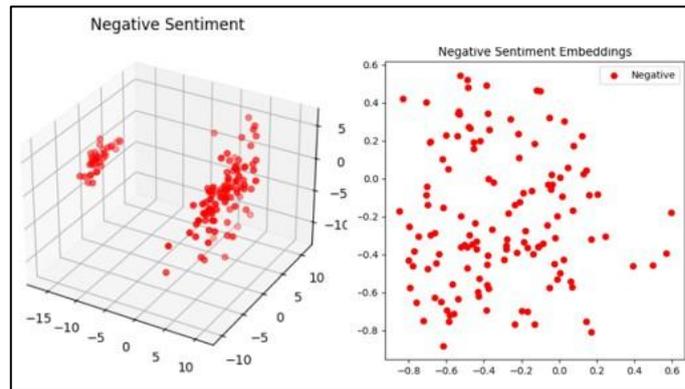
### 3.1.3. Embedding Data

*Embedding* ini akan memetakan teks ke ruang vektor, di mana model dapat menangkap perbedaan konteks antara teks dengan sentimen positif, netral dan negatif. Berdasarkan distribusi data sentimen positif : 13,872, sentimen netral : 642, dan sentimen negatif : 131 maka jumlah data sentimen positif jauh lebih tinggi dibandingkan netral dan negatif, yang dapat memengaruhi analisis visual dan hasil pemodelan.

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Sentiment
0	-0.419348	0.20028	-0.585158	-0.998595	1	Negative
1	0.122664	0.224798	0.303495	-0.658278	0.998402	Negative
2	0.0250486	-0.425269	-0.104888	-0.989906	0.999789	Negative
3	-0.44168	-0.208719	0.119917	-0.83455	0.999976	Neutral
4	-0.199717	-0.427499	-0.0876511	0.309063	0.999028	Neutral
5	-0.694328	0.119489	-0.448113	-0.999523	0.999999	Neutral
6	-0.203173	0.651678	-0.489809	0.995967	0.602871	Positive
7	-0.555077	0.704188	-0.330412	-0.931661	0.956916	Positive
8	-0.661834	0.544007	-0.123002	0.410503	0.679046	Positive

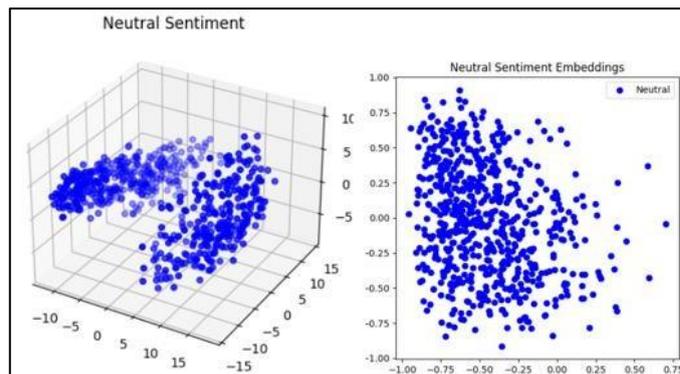
Gambar 8. Dataframe hasil embedding

*Scatter plot* yang dibuat menunjukkan tiga kategori sentimen dalam bentuk sebaran titik pada ruang vektor dua dimensi. Sumbu x dan y pada masing-masing plot merepresentasikan dimensi-dimensi hasil reduksi dari *embedding* BERT, sehingga kita bisa melihat perbedaan distribusi berdasarkan kategori sentimen. Plot pada Gambar 99 menunjukkan distribusi *embedding* untuk teks dengan sentimen negatif. Titik-titik yang tersebar di seluruh area menandakan variasi *embedding* yang cukup besar dalam teks dengan sentimen negatif, meskipun jumlahnya paling sedikit yaitu 131 data, penyebaran ini menunjukkan bahwa model mungkin menangkap berbagai ekspresi negatif, namun karena jumlahnya sedikit, pola distribusinya tidak terlalu padat.



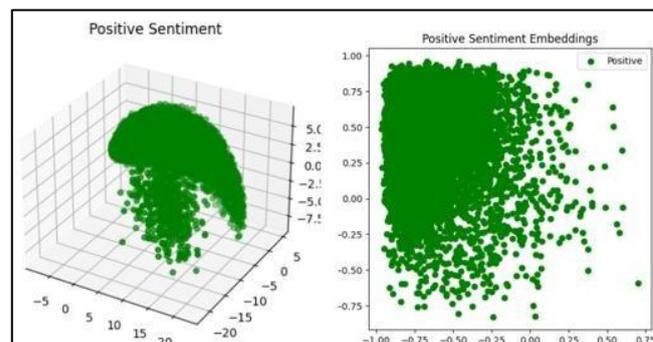
Gambar 9. Scatter plot negatif sentimen

Plot pada Gambar 10 menunjukkan distribusi *embedding* untuk teks netral. Titik biru tersebar di sekitar area tertentu, dengan cakupan yang sedikit lebih luas daripada sentimen negatif. Dengan jumlah 642 data, distribusi ini terlihat cukup menyebar dan tidak terlalu padat, menunjukkan bahwa sentimen netral mungkin memiliki variasi dalam pola bahasa yang ditangkap oleh *embedding*.



Gambar 10. Scatter plot netral sentimen

Plot pada Gambar 11 memperlihatkan *embedding* untuk sentimen positif, yang memiliki jumlah sampel paling banyak yaitu 13,872 data. Distribusi titik hijau yang sangat padat di area tertentu menunjukkan bahwa *embedding* untuk sentimen positif memiliki pola yang lebih homogen dibandingkan sentimen lain. Kepadatan ini menunjukkan bahwa model mungkin lebih mampu mengidentifikasi pola-pola tertentu yang umum dalam teks positif, sehingga terbentuk kluster yang lebih terkonsentrasi.



Gambar 11. Scatter plot positif sentimen

Distribusi ini juga menunjukkan bahwa *embedding* BERT berhasil mengelompokkan teks ke dalam kluster yang cukup terpisah untuk masing-masing sentimen, Sentimen negatif dan netral memiliki distribusi yang lebih menyebar, menunjukkan bahwa mereka mencakup beragam pola bahasa yang mungkin sulit ditangkap oleh model dengan *dataset* saat ini.

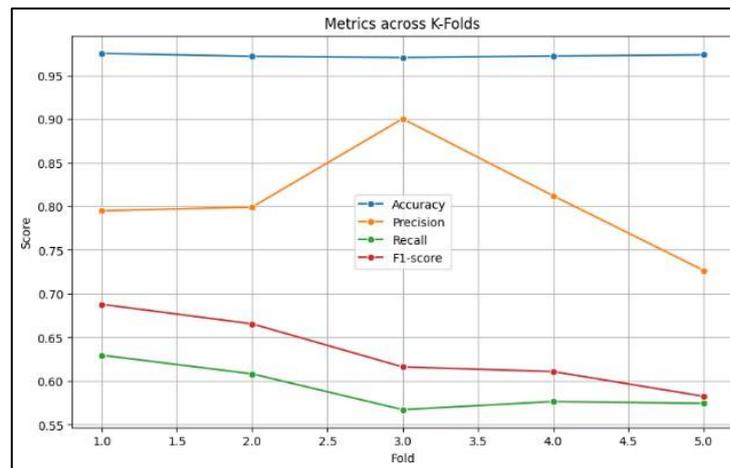
### 3.1.4. Pelatihan Model SVM

Pada tahapan ini data teks yang diubah menjadi representasi numerik menggunakan *embedding* dari model BERT (*indobert-large-p1*), kemudian digunakan sebagai fitur *input* ke model *Support Vector Machine* (SVM). Hasil evaluasi dari pelatihan model SVM sendiri adalah Negatif (*Precision: 0.57 Recall: 0.36 dan f1-score: 0.44*), Netral (*Precision: 0.81 Recall: 0.62 f1-score: 0.70*), dan Positif (*Precision: 0.98 Recall: 1.00 f1-score: 0.99*) Akurasi Keseluruhan pada model ini adalah: 0.97.

Evaluating SVM model...				
	precision	recall	f1-score	support
Negative	0.57	0.36	0.44	11
Neutral	0.81	0.62	0.70	69
Positive	0.98	1.00	0.99	1385
accuracy			0.97	1465
macro avg	0.79	0.66	0.71	1465
weighted avg	0.97	0.97	0.97	1465

Gambar 12. Evaluasi Model SVM

Pada tahapan evaluasi model SVM dengan *embedding* BERT menggunakan *k-fold cross-validation*, didapatkan hasil metrik rata-rata di 5 *folds* *Accuracy: 0.9725 Precision: 0.8066 Recall: 0.5911 F1-score: 0.6325*



Gambar 13. Metrik k-folds

Pada Gambar 13. *accuracy* pada setiap *fold* menunjukkan nilai yang sangat tinggi (di atas 95%) dan konsisten di semua *fold*. *precision* meningkat signifikan pada *fold* ke-3 sebelum menurun pada *fold* berikutnya, *F1-score*, yang merupakan rata-rata harmonis antara *precision* dan *recall*, juga menunjukkan tren menurun seiring bertambahnya *fold* dapat dilihat tren *precision* yang lebih tinggi dibandingkan *recall* menunjukkan bahwa model lebih fokus pada kelas yang mudah diprediksi (kelas mayoritas).

### 3.2. Pembahasan

Pembersihan data menggunakan pendekatan berbasis kemiripan teks telah berhasil menghapus sejumlah data yang dianggap redundant (duplikat) dengan tetap mempertahankan integritas informasi utama. Setelah pembersihan, terjadi sedikit perubahan pada rata-rata interaksi di metrik favorit, *quote*, *reply*, dan *retweet*. Selanjutnya hasil inferensi awal menunjukkan adanya distribusi data yang lebih dominan pada sentimen positif, model dapat cenderung lebih akurat dalam mengidentifikasi sentimen positif, tetapi kurang dalam mengenali sentimen negatif dan netral. Penyebaran yang lebih luas untuk kategori negatif dan netral menandakan bahwa terdapat lebih banyak variasi dalam jenis teks yang termasuk dalam kategori ini. Keberagaman dalam distribusi sentimen negatif dan netral juga disebabkan oleh rendahnya jumlah data pada kedua kategori ini, yang mana hal ini berpotensi memengaruhi kualitas klasifikasi model.

Adanya ketidakseimbangan kelas sangat jelas terlihat dalam data pengujian, hal ini mengakibatkan redahnya nilai *recall* dan *f1-score* pada kelas negatif

Berdasarkan hasil evaluasi, hasil negatif *Precision*: 0.57 artinya, dari semua prediksi negatif, 57% benar-benar negatif. *Recall*: 0.36 artinya, dari semua data yang sebenarnya negatif, model mendeteksi 36% sebagai negatif. *f1-score*: 0.44 artinya, nilai *f1-score* yang relatif rendah menunjukkan bahwa model kesulitan dalam mengidentifikasi kelas negatif, kemungkinan besar karena jumlah data negatif yang sedikit.

Netral *Precision*: 0.81 artinya, dari semua prediksi netral, 81% benar-benar netral. *Recall*: 0.62 artinya, dari semua data yang sebenarnya netral, 62% terdeteksi dengan benar. *f1-score*: 0.70 artinya, kinerja pada kelas netral cukup baik, meskipun masih ada ruang untuk perbaikan dalam hal *recall*.

Positif *Precision*: 0.98 artinya, hampir semua prediksi positif benar, menunjukkan bahwa model sangat baik dalam mendeteksi sentimen positif. *Recall*: 1.00 artinya, semua data yang sebenarnya positif diklasifikasikan dengan benar sebagai positif. *f1-score*: 0.99 artinya, *f1-score* yang mendekati sempurna menunjukkan bahwa model sangat efektif dalam mengidentifikasi kelas positif, yang dominan dalam *dataset*.

Akurasi Keseluruhan: 0.97 artinya, model memiliki akurasi keseluruhan yang sangat tinggi, yaitu 97%. Ini mengindikasikan bahwa secara umum model dapat mengklasifikasikan sentimen dengan baik pada data pengujian.

*Macro Average*: nilai rata-rata dari metrik (*precision*, *recall*, *f1-score*) di antara semua kelas, tanpa mempertimbangkan frekuensi kelas. Pada *macro average*, *f1-score* adalah 0.71, yang menunjukkan bahwa performa model tidak seimbang antar kelas.

*Weighted Average*: nilai rata-rata dari metrik, di mana masing-masing kelas memiliki bobot berdasarkan jumlah datanya. Dengan *weighted f1-score* 0.97, dari sini kita bisa melihat bahwa secara keseluruhan model bekerja sangat baik, tetapi hal cenderung terpengaruh oleh kelas positif yang dominan.

Hasil evaluasi model dengan menggunakan *k-fold cross-validation* menunjukkan *precision* terlihat tinggi pada *fold* tertentu hal ini memperjelas bahwa model lebih akurat dalam menghindari *false positive* (prediksi salah pada kelas mayoritas).

## 4. DISCUSSIONS

Dalam implementasinya proses pembersihan data menggunakan pendekatan berbasis kemiripan teks telah berhasil menghapus sejumlah data yang dianggap redundant (duplikat) dengan tetap mempertahankan integritas informasi utama. Setelah pembersihan, terjadi sedikit perubahan pada rata-rata interaksi di metrik favorit, *quote*, *reply*, dan *retweet*. Selanjutnya hasil inferensi awal menunjukkan adanya distribusi data yang lebih dominan pada sentimen positif, model yang di hasilkan dapat cenderung lebih akurat dalam mengidentifikasi sentimen positif, tetapi kurang dalam mengenali

sentimen negatif dan netral. Penyebaran yang lebih luas untuk kategori negatif dan netral menandakan bahwa terdapat lebih banyak variasi dalam jenis teks yang termasuk dalam kategori tersebut. Keberagaman dalam distribusi sentimen negatif dan netral juga disebabkan oleh rendahnya jumlah data pada kedua kategori ini, yang mana hal ini berpotensi memengaruhi kualitas klasifikasi model, sehingga pada penelitian ini tidak dapat secara efektif mengidentifikasi sentimen negatif dan netral, hal ini disebabkan karena adanya ketidakseimbangan pada kelas, imbas dari ketidakseimbangan tersebut mengakibatkan rendahnya nilai *recall* dan *f1-score* pada kelas negatif.

Proses pembersihan data berhasil mengurangi data duplikat, tetapi distribusi sentimen tetap didominasi oleh sentimen positif. Meskipun integrasi BERT dan SVM terbukti efektif dalam menangkap nuansa sentimen, tantangan dalam mencapai kinerja yang lebih seimbang di semua kategori sentimen masih ada.

Dari hasil distribusi sentimen pada Gambar 7 dapat dilihat dengan jelas adanya ketidakseimbangan pada kelas-kelas yang telah dihasilkan adapun penelitian terdahulu pernah mencoba melakukan penggabungan data pada kelas minoritas akan tetapi tindakan tersebut tidak berhasil[27], juga pada penelitian sebelumnya pernah dilakukan *oversampling* dengan pendekatan gabungan SMOTE (*Synthetic Minority Oversampling Technique*) dan PSO (*Particle Swarm Optimization*), akan tetapi cara ini juga kurang tepat untuk menhandel data teks[28]

Pada penelitian sebelumnya telah dilakukan juga penambahan *dataset* untuk melakukan penyeimbangan data sehingga mendapatkan nilai akurasi yang cukup tinggi sebesar 96%,[16] sedangkan pendekatan yang dilakukan pada penelitian ini menghasilkan akurasi yang lebih baik dengan data yang tidak seimbang.

## 5. CONCLUSION

Pada proses data *preprocessing* terdapat peningkatan dan penurunan pada metrik interaksi (*favorite, quote, reply, retweet*) menunjukkan bahwa pada proses pembersihan ini memiliki dampak pada distribusi interaksi antar *tweet*, namun tanpa perubahan yang signifikan pada nilai-nilai rata-ratanya.

Dari hasil distribusi sentimen, terlihat adanya dominasi sentimen positif, dengan netral dan negatif yang jauh lebih sedikit. Ini karena data yang digunakan memang cenderung mengarah pada sentimen positif. Model BERT sendiri memang lebih cenderung memberi label positif pada teks yang ambigu.

Distribusi *scatter plot* pada Gambar 9, Gambar 10, Gambar 11 memberikan gambaran tentang bagaimana *embedding* BERT memetakan teks ke ruang vektor berdasarkan kategori sentimen. Sentimen positif lebih padat dalam area tertentu, sementara negatif dan netral menunjukkan distribusi yang lebih menyebar. Akurasi tinggi yang ditunjukkan oleh model dipengaruhi oleh jumlah besar data positif. Model menghasilkan kinerja yang baik untuk kelas positif, kinerja pada kelas negatif masih perlu diperbaiki. Performa yang kurang optimal pada kelas negatif dan netral mengindikasikan bahwa model mengalami ketidakseimbangan kelas. Model lebih fokus pada deteksi kelas mayoritas (positif), sehingga lebih sulit untuk mengenali sentimen negatif dan netral. Integrasi BERT Dan SVM dalam analisis sentimen berhasil dan dapat dilakukan, terlihat juga model ini mampu menangkap nuansa sentimen dengan baik.

Hasil utama penelitian menunjukkan bahwa integrasi ini meningkatkan akurasi klasifikasi sentimen secara keseluruhan hingga 97%. Namun, terdapat ketidakseimbangan kelas yang mempengaruhi akurasi sentimen negatif dan netral. Meskipun pendekatan gabungan ini efektif untuk klasifikasi sentimen, diperlukan perbaikan lebih lanjut untuk mencapai kinerja yang lebih seimbang di semua kategori sentimen.

Untuk meningkatkan performa model dan mengatasi ketidakseimbangan kelas dalam klasifikasi sentimen, disarankan untuk memanfaatkan *transfer learning* dengan menggunakan model *pretrained* yang telah dilatih pada *dataset* yang lebih besar dan seimbang. Hal ini dapat membantu model memahami konteks yang lebih luas dan meningkatkan generalisasi terhadap data baru. Selain itu, penggunaan model generatif seperti GPT (*Generative Pre-trained Transformer*) untuk menghasilkan teks sintesis yang berkualitas dapat menjadi solusi dalam melakukan *oversampling* pada kelas minoritas, sehingga distribusi data menjadi lebih seimbang tanpa mengorbankan keragaman data. Selanjutnya, integrasi analisis aspek (*aspect-based sentiment analysis*) juga direkomendasikan untuk menangkap nuansa sentimen yang lebih detail pada aspek tertentu dalam teks, yang dapat memberikan wawasan lebih mendalam tentang interaksi antar-sentimen di berbagai konteks.

## CONFLICT OF INTEREST

Dengan ini saya menyatakan bahwa penelitian ini tidak ada unsur konflik kepentingan dari berbagai pihak manapun, baik dari pihak IIB Darmajaya maupun penulis yang tercantum pada artikel ini. Segala bentuk pembiayaan berasal dari dana pribadi.

## ACKNOWLEDGEMENT

Terimakasih kepada Halimah yang telah membantu dalam proses perhitungan secara manual, Sutedi yang telah membantu dalam mencari referensi, Melda Aghariana yang telah membantu menyempurnakan hasil tulisan ini dan M. Reza Redo Islami yang telah membantu mencari data untuk penelitian ini.

## REFERENCES

- [1] Perludem, "Putusan MK Nomor 60/PUU-XXII/2024 Tentang Ambang Batas Pencalonan Kepala Daerah di Pilkada," Perkumpulan untuk Pemilu dan Demokrasi. Accessed: Oct. 03, 2024. [Online]. Available: <https://perludem.org/2024/08/28/putusan-mk-nomor-60-puu-xxii-2024/>
- [2] R. Redo, *Data Tweet 20-Agustus-2024 s/d 27-Agustus-2024 (key:Putusan MK-No.60/PUU-XXII/2024)*. 2024. doi: 10.13140/RG.2.2.32656.42242.
- [3] M. Agarina, S. Sutedi, A. S. Karim, and E. R. Sari, "Strategi Marketing Promosi Penerimaan Mahasiswa Baru Menggunakan K-Means Clustering," *TEKNIKA 18 (2)*, 673–682, vol. 682. 2024.
- [4] D. Wijaya and S. Sutedi, "Analisis Sentimen Opini Masyarakat Terhadap Penerapan Pelayanan Digital Di Desa Margo Mulyo Menggunakan Algoritma K-Nearest Neighbor," *JURNAL ILMU KOMPUTER, SISTEM INFORMASI, TEKNIK INFORMATIKA*, vol. 3, no. 1. pp. 13–26, 2024.
- [5] M. E. M. Abo, "A multi-criteria approach for arabic dialect sentiment analysis for online reviews: Exploiting optimal machine learning algorithm selection," *Sustainability (Switzerland)*, vol. 13, no. 18. 2021. doi: 10.3390/su131810018.
- [6] H. Adamu, "Framing twitter public sentiment on Nigerian government COVID-19 palliatives distribution using machine learning," *Sustainability (Switzerland)*, vol. 13, no. 6. 2021. doi: 10.3390/su13063497.
- [7] H. Kurniawan, J. Triloka, and Y. Ardhan, "Analysis of the Artificial Neural Network Approach in the Extreme Learning Machine Method for Mining Sales Forecasting Development," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1. 2023.
- [8] H. Nisar and Wasilah, "Pemanfaatan K Means Clustering dalam Pengelompokan Judul Skripsi," *Jurnal Penelitian Ilmu Dan Teknologi Komputer*, vol. 14, no. 1. pp. 19–26, 2022.
- [9] D. M. Abdullah, "Machine Learning Applications based on SVM Classification: A Review," *Qubahan Academic Journal*, vol. 1, no. 2. pp. 81–90, 2021. doi: 10.48161/qaj.v1n2a50.

- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." 2019. [Online]. Available: <https://arxiv.org/abs/1810.04805>
- [11] N. Chintalapudi, "Sentimental analysis of COVID-19 tweets using deep learning models," *Infectious Disease Reports*, vol. 13, no. 2. 2021. doi: 10.3390/IDR13020032.
- [12] Y. Cui, "Pre-Training with Whole Word Masking for Chinese BERT," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 29. pp. 3504–3514, 2021. doi: 10.1109/TASLP.2021.3124365.
- [13] S. Bengesi, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," *IEEE Access*, vol. 11. pp. 11811–11826, 2023. doi: 10.1109/ACCESS.2023.3242290.
- [14] Z. Zhang, "A Study on Early Warnings of Financial Crisis of Chinese Listed Companies Based on DEA-SVM Model," *Mathematics*, vol. 10, no. 12. 2022. doi: 10.3390/math10122142.
- [15] S. Y. Irianto, S. Karnila, and D. Yuliawati, "Study of Manhattan and Region Growing Methods for Brain Tumor Detection," *Journal of Advances in Information Technology*, vol. 15, no. 2. 2024.
- [16] G. Bourahouat, "Improvement of Moroccan Dialect Sentiment Analysis Using Arabic BERT-Based Models," *Journal of Computer Science*, vol. 20, no. 2. pp. 157–167, 2024. doi: 10.3844/jcssp.2024.157.167.
- [17] Fransiscus, "Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method," *International Journal of Engineering Trends and Technology*, vol. 70, no. 12. pp. 281–288, 2022. doi: 10.14445/22315381/IJETT-V70I12P226.
- [18] M. Mozafari, "Hate speech detection and racial bias mitigation in social media based on BERT model," *PLoS ONE*, vol. 15, no. 8. 2020. doi: 10.1371/journal.pone.0237861.
- [19] T. Wang, "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model," *IEEE Access*, vol. 8. pp. 138162–138169, 2020. doi: 10.1109/ACCESS.2020.3012595.
- [20] L. Zhao, "A BERT based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts," *Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2021*. pp. 1233–1238, 2021. doi: 10.1109/CSCWD49262.2021.9437616.
- [21] M. Singh, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Social Network Analysis and Mining*, vol. 11, no. 1. 2021. doi: 10.1007/s13278-021-00737-z.
- [22] R. K. Kaliyar, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 8. pp. 11765–11788, 2021. doi: 10.1007/s11042-020-10183-2.
- [23] L. K. Ramasamy, "Performance analysis of sentiments in Twitter dataset using SVM models," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3. pp. 2275–2284, 2021. doi: 10.11591/ijece.v11i3.pp2275-2284.
- [24] K. X. Han, "Application of support vector machine (SVM) in the sentiment analysis of twitter dataset," *Applied Sciences (Switzerland)*, vol. 10, no. 3. 2020. doi: 10.3390/app10031125.
- [25] K. Aurangzeb, "Aspect Based Multi-Labeling Using SVM Based Ensembler," *IEEE Access*, vol. 9. pp. 26026–26040, 2021. doi: 10.1109/ACCESS.2021.3055768.
- [26] "config.json · indobenchmark/indobert-base-p1 at main." Accessed: Nov. 30, 2024. [Online]. Available: <https://huggingface.co/indobenchmark/indobert-base-p1/blob/main/config.json>
- [27] M. R. Redo and A. Perdana, "ANALISIS KETIDAKSEIMBANGAN KELAS DALAM PENGEMBANGAN MODEL KLASIFIKASI," *Proceedings of the National Conference on Electrical Engineering .... 2023*.
- [28] R. Obiedat, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10. pp. 22260–22273, 2022. doi: 10.1109/ACCESS.2022.3149482.

