

Sentiment Analysis of Public Tendencies Towards The Israel-palestine Conflict Using Bert Transformer on Platform X

[Sentimen Analisis Tendensi Masyarakat terhadap Konflik Israel-palestina menggunakan Transformer Bert Di Platfrom X]

Syaiful Mulki Almubarok Renhoran¹⁾, Hamzah Setiawan²⁾

¹⁾ Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: hamzah@umsida.ac.id

Abstract. *This research addresses the analysis of public sentiment towards the Israel-Palestine conflict using the BERT Transformer model on the X platform. The main issues include the effectiveness of NLP methods for analyzing positive, negative, and neutral sentiments and the accuracy of these methods. The study implements a sentiment analysis algorithm using BERT, achieving an accuracy of 93% with precision, recall, and F1-Score of 0.95, 0.93, and 0.94, respectively. These results indicate high performance of the sentiment classification model. The conclusion highlights the model's satisfactory performance in predicting sentiment from tweets, although there is room for improvement, especially in handling neutral sentiments. This model can be confidently used for social media analysis, brand monitoring, and reputation management.*

Keywords - Sentiment analysis; BERT Base-Uncase; Bidirectional Encoder Representations from Transformer (BERT); Natural Language Processing (NLP); Tendencies on Platform X

Abstrak. *Penelitian ini membahas analisis sentimen publik terhadap konflik Israel-Palestina menggunakan model Transformer BERT di platform X. Masalah utama meliputi efektivitas metode NLP untuk menganalisis sentimen positif, negatif, dan netral serta akurasi metode tersebut. Studi ini mengimplementasikan algoritma analisis sentimen menggunakan BERT, mencapai akurasi sebesar 93% dengan precision, recall, dan F1-Score masing-masing sebesar 0,95, 0,93, dan 0,94. Hasil ini menunjukkan kinerja tinggi dari model klasifikasi sentimen. Kesimpulan menyoroti kinerja memuaskan model dalam memprediksi sentimen dari tweet, meskipun masih ada ruang untuk perbaikan, terutama dalam menangani sentimen netral. Model ini dapat digunakan dengan percaya diri untuk analisis media sosial, pemantauan merek, dan pengelolaan reputasi.*

Kata Kunci - Analisis sentimen; BERT Base-Uncase; Bidirectional Encoder Representations from Transformer (BERT); Natural Language Processing (NLP); Tendensi pada Platform X

I. PENDAHULUAN

Konflik antara Palestina dan Israel yang dipicu oleh serangan Hamas terhadap Israel di Jalur Gaza pada 7 Oktober telah menimbulkan dampak signifikan, baik secara geopolitik maupun ekonomi global. Dengan lebih dari 4.900 korban jiwa dan lebih dari 25.000 orang terluka, konflik ini menarik perhatian internasional dan memicu respons dari berbagai negara, termasuk evakuasi warga negara asing dari wilayah konflik.

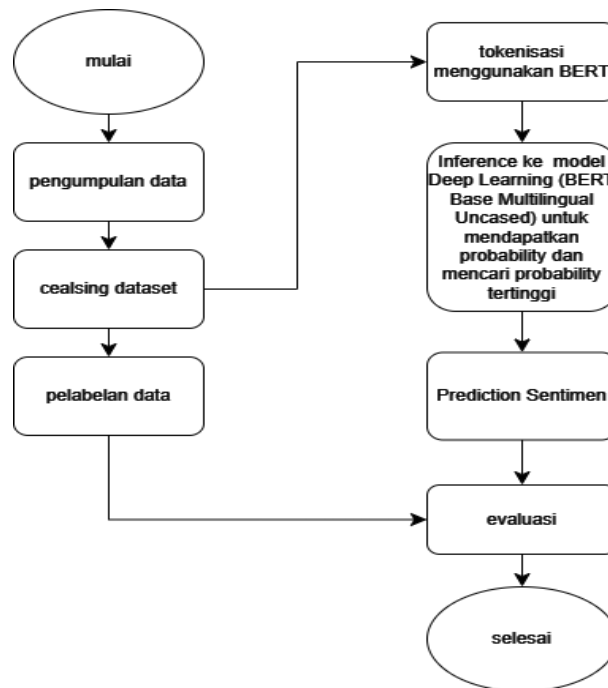
Penelitian ini bertujuan untuk menganalisis tendensi masyarakat Indonesia terhadap konflik Palestina-Israel melalui media sosial X. Metode yang digunakan adalah Deep Learning dengan arsitektur model BERT (Bidirectional Encoder Representations from Transformers) yang dikembangkan oleh Google. BERT memungkinkan pemahaman konteks dan hubungan antar kata dalam teks, yang berguna untuk analisis sentimen positif, negatif, dan netral.

Tujuan utama penelitian ini adalah mengimplementasikan algoritma NLP (Natural Language Processing) untuk mengidentifikasi sentimen publik dan mengevaluasi akurasi. Batasan penelitian meliputi fokus pada pengguna media sosial X di Indonesia dan analisis tweet berbahasa Indonesia terkait konflik ini. Data dikumpulkan dari 1 Mei 2024 hingga saat ini.

Manfaat penelitian ini termasuk memberikan wawasan mendalam tentang dinamika opini publik, mendukung upaya diplomatik, dan membantu dalam pengelolaan reputasi di media sosial. Hasil analisis diharapkan dapat membantu dalam pencegahan eskalasi konflik melalui deteksi potensi ketegangan dan memberikan panduan untuk kampanye informasi yang terarah.

II. METODE

Gambar 1. merupakan alur diagram sistem yang diterapkan pada penelitian ini.



Gambar 1. Alur diagram sistem

A. Pengumpulan data

Data yang digunakan dalam penelitian ini berasal dari tweet “KONFLIK ISRAEL-PALESTINA” berbahasa indonesia yang diambil dari platfrom X. Data diambil dari tanggal 01 Mei - sekarang Untuk mendapatkan data dilakukan Crawl Data dengan menggunakan Google Colab. Pertama, diinstal library pandas untuk memudahkan manipulasi data setelah diambil. Selanjutnya, kita perlu menginstal Node.js karena tweet-harvest dibangun menggunakan Node.js . Langkah instalasi ini melibatkan beberapa perintah shell untuk update paket, instalasi dependensi, dan mengatur konfigurasi agar dapat mengambil kunci dari sumber resmi Node.js. Setelah Node.js berhasil diinstal, kita dapat menggunakan perintah tweet-harvest untuk melakukan crawling data dari Twitter sesuai dengan kata kunci pencarian yang ditentukan. Hasil crawling akan disimpan dalam format CSV dengan nama file yang sudah ditentukan sebelumnya. Selanjutnya, dengan menggunakan pandas, kita membaca file CSV tersebut ke dalam sebuah DataFrame untuk analisis lebih lanjut. Dengan langkah-langkah ini, proses scraping data dari Twitter dapat dilakukan secara efisien di Google Colab.

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
created_at	full_text	id_str	page_to_screen	lang	lat	lon	retweet_count	reply_count	retweet_id	retweet_username	tweet_url	user_id	username	Cleaned_Tweet
Tue May	0	Pengungs	1,79037E	https://pbs.twimg	in		0	0	0	0	https://tv	3857319E	vthreeizm	pengungsian paksa warga palestina dari rafah
Tue May	0	Pemukim	1,79037E	https://pbs.twimg	in	li	0	0	0	0	https://tv	1,11886E	Parstoday	pemukim zionis mencegah dan menyerang truk yang membawa bantuan medis untuk warga palestina
Tue May	0	@_khana	1,79037E	_khanaza	in		0	0	0	0	https://tv	30321521	lucimaya1	tidak seorang pun dari zionist yahudi laknatullah tidak berdosa bahkan anakanak pun telah menangg
Tue May	1	Jangan se	1,79037E		in	lo	0	0	0	0	https://tv	3659355E	fallfortaej	jangan sembarangan ngeluarin opini kalo km pake bendera palestina di dn mu tadi aku lihat udh ada y
Tue May	0	@CNNInc	1,79037E	CNNIndor	in	Re	0	0	0	0	https://tv	7,08706E	peppew6i	jangan pernah lelah bicara tentang tanah palestina
Tue May	0	Biadab! Is	1,79037E		in		0	0	0	0	https://tv	1,22543E	RepelitaO	biadab israel sengaja bunuh para ulama untuk lemahkan iman rakyat palestina
Tue May	0	@convomf	1,79037E	convomfs	in	EX	0	0	0	0	https://tv	7,32105E	loeycrush	dia terangterangan nyebut gaza semoga yang lain juga bisa speak up untuk bela palestina
Tue May	0	gila ya ter	1,79037E		in	fa	0	1	0	0	https://tv	1,08037E	dexkthv	gila ya ternyata orang tt banyak yang kurang empati banget abt palestina dan produk yang diboikot di
Tue May	0	sakit bgt	1,79037E		in	Dc	0	0	0	0	https://tv	2426520C	radiaham	sakit bgt setiap hari liat orang orang palestina ya allah bantuin mereka ya allah

Gambar 2. Scrapping Data

B. Cleaning Data

Tahap cleaning data akan membuang karakter yang tidak memiliki makna dan membuat data menjadi lebih konsisten. Pada penelitian ini cleaning data dilakukan dengan menghapus tag HTML, case folding, menghapus angka, tanda baca, dan karakter

P
Cleaned_Tweet
pengungsian paksa warga palestina dari rafah
pemukim zionis mencegat dan menyerang truk yang membawa bantuan medis untuk warga palestina di jalur gaza mereka menghancurkan muatan truk ini
tidak seorang pun dari zionist yahudi laknatullah tidak berdosa bahkan anakanak pun telah menanggung dosanya sendiri dengan menanamkan rasa benci iri hati tidak berperikemanusiaan terhadap palestina se
jangan sembarangan ngeluarin opini kalo km pake bendera palestina di dn mu tadi aku lihat udh ada yg edukasi km ya kalo udh paham tolong hapus tweetnya sebelum km jd miss_muslim pt 2 kalo ga mau hapus
jangan pernah lelah bicara tentang tanah palestina
biadab israel sengaja bunuh para ulama untuk lemahkan iman rakyat palestina
dia terangterangan nyebut gaza semoga yang lain juga bisa speak up untuk bela palestina
gila ya ternyata orang tt banyak yang kurang empati banget abt palestina dan produk yang diboikot di indo masih banyak yang ga waras kek ga mikir juga gitu masa diminta jgn beli produk md aj malah bandingin
sakit bgt setiap hari liat orang orang palestina ya allah bantuin mereka ya allah
lu percaya itu fotonya di palestina gaza bisa aja itu cuma fotovideo demo utk di europa
iya tau emg mengecewakan tp skrg bkn soal artis2 lg kalian juga hrs gencar skrg share tentang palestina menurut gue hipokrit juga klo ngarepin artis udh banyak artis yg signed buat ceasefire tp sampe skrg ga be
allaahu akbaryaa allah limpahkanlah pertolongan perlindungan bantuan kekuatan kemulyaan lahir batin dunia akhirat bagi saudarasaudara kami muslimin muslimat di palestina aamiiiiin
nois p krl
breaking news seruan boikot penjajah israel suasana aksi demonstrasi di ibukota tokyo jepang pada ahad 1252024 aksi tersebut bertujuan untuk mendukung palestina dan menolak genosida di jalur gaza
ya allah bebaskan negeri palestina dr kekejaman orgng dzolim ya allah lindungilah seluruh warga palestina dr marahlbahaya dan merdekakan mereka semua ya allah serta musnahkan dr bumi ini orang orang
mendukung palestina dgn menyebar hoax bahlul ente
wkwkwkwk jadi kami dari dunia belahan lain ini kudu ngapain goblok aku udah donasi udah ikut gebrakan palestina di kotaku kemarin sedangkan kau cuma gunain palestina buat stan war parah memang
raja abdullah mengutuk rencana israel untuk menyerang rafah menyerukan gencatan senjata dan pembentukan negara palestina Biden menjamu raja abdullah yordania di gedung putih senindi sambutan publik
alhamdulillah
asikkkkk mancing keributan babawaba palestina
perlu dipertahankan begitu pula halnya dengan jabalia diserang sepenuhnya dan sepenuhnya pejuang palestina diserang dari arah selatan lingkungan al zaytoun dan israel menyerang mereka untuk memastikan
mereka di gaza utara israel sedang beroperasi penuh di sini sekarang serangan udara dan artileri secara bersamaan terjadi di sana lingkungan al zaytoun juga penting bagi para pejuang palestina karena merupakan

Gambar 3. Cleaning Data

C. Pelabelan Data

Pelabelan dilakukan oleh peneliti secara manual. Pada setiap ulasan akan diberikan label yaitu label sentimen. Untuk label sentimen terbagi menjadi tiga yaitu positif, negatif, dan netral. Pemilihan aspek dilakukan dengan melihat tweet yang sering dibahas oleh para pengguna

P	Q
Cleaned_Tweet	Sentiment
pengungsian paksa warga palestina dari rafah	negatif
pemukim zionis mencegat dan menyerang truk yang membawa bantuan medis untuk warga palestina di jalur gaza mereka menghancurkan muatan truk ini	negatif
tidak seorang pun dari zionist yahudi laknatullah tidak berdosa bahkan anakanak pun telah menanggung dosanya sendiri dengan menanamkan rasa benci iri hati tidak berperikemanusiaan terhadap	netral
jangan sembarangan ngeluarin opini kalo km pake bendera palestina di dn mu tadi aku lihat udh ada yg edukasi km ya kalo udh paham tolong hapus tweetnya sebelum km jd miss_muslim pt 2 kalo ga	netral
jangan pernah lelah bicara tentang tanah palestina	positif
biadab israel sengaja bunuh para ulama untuk lemahkan iman rakyat palestina	netral
dia terangterangan nyebut gaza semoga yang lain juga bisa speak up untuk bela palestina	positif
gila ya ternyata orang tt banyak yang kurang empati banget abt palestina dan produk yang diboikot di indo masih banyak yang ga waras kek ga mikir juga gitu masa diminta jgn beli produk md aj mala	netral
sakit bgt setiap hari liat orang orang palestina ya allah bantuin mereka ya allah	positif
lu percaya itu fotonya di palestina gaza bisa aja itu cuma fotovideo demo utk di europa	negatif
iya tau emg mengecewakan tp skrg bkn soal artis2 lg kalian juga hrs gencar skrg share tentang palestina menurut gue hipokrit juga klo ngarepin artis udh banyak artis yg signed buat ceasefire tp sam	positif
allaahu akbaryaa allah limpahkanlah pertolongan perlindungan bantuan kekuatan kemulyaan lahir batin dunia akhirat bagi saudarasaudara kami muslimin muslimat di palestina aamiiiiin	positif
nois p krl	negatif
breaking news seruan boikot penjajah israel suasana aksi demonstrasi di ibukota tokyo jepang pada ahad 1252024 aksi tersebut bertujuan untuk mendukung palestina dan menolak genosida di jalur	positif
ya allah bebaskan negeri palestina dr kekejaman orgng dzolim ya allah lindungilah seluruh warga palestina dr marahlbahaya dan merdekakan mereka semua ya allah serta musnahkan dr bumi ini	positif
mendukung palestina dgn menyebar hoax bahlul ente	positif
wkwkwkwk jadi kami dari dunia belahan lain ini kudu ngapain goblok aku udah donasi udah ikut gebrakan palestina di kotaku kemarin sedangkan kau cuma gunain palestina buat stan war parah me	negatif
raja abdullah mengutuk rencana israel untuk menyerang rafah menyerukan gencatan senjata dan pembentukan negara palestina Biden menjamu raja abdullah yordania di gedung putih senindi samb	netral
alhamdulillah	positif
asikkkkk mancing keributan babawaba palestina	negatif
perlu dipertahankan begitu pula halnya dengan jabalia diserang sepenuhnya dan sepenuhnya pejuang palestina diserang dari arah selatan lingkungan al zaytoun dan israel menyerang mereka untuk r	positif
mereka di gaza utara israel sedang beroperasi penuh di sini sekarang serangan udara dan artileri secara bersamaan terjadi di sana lingkungan al zaytoun juga penting bagi para pejuang palestina karer	netral

Gambar 4. Pelabelan Data

D. Implementasi BERT

Sebelum proses pelatihan dimulai, data yang telah terkumpul perlu diubah menjadi format input yang dapat diterima oleh BERT. Untuk mencapai hal ini, digunakan proses tokenisasi yang bertujuan untuk mengonversi setiap kata ke dalam token-token yang dapat dipahami oleh model

1. Pengumpulan Data

Pengguna (misalnya, peneliti atau praktisi data) memulai dengan mengumpulkan data ulasan dari platform X. Platform X dapat berupa situs web atau aplikasi yang menyediakan ulasan dari berbagai pengguna. Langkah ini melibatkan ekstraksi teks ulasan dari platform tersebut yang kemudian akan digunakan untuk analisis lebih lanjut

2. Pembersihan Data

Setelah data ulasan terkumpul, data tersebut dikirim ke tahap pembersihan data. Di sini, data dibersihkan dari karakter yang tidak relevan, angka, tanda baca, dan karakter khusus yang tidak diperlukan untuk analisis. Langkah ini penting untuk memastikan bahwa teks ulasan dalam kondisi yang siap untuk diproses lebih lanjut tanpa gangguan dari elemen-elemen yang tidak relevan.

3. Preprocessing Data

Data yang telah dibersihkan kemudian dikirim ke tahap preprocessing. Preprocessing mencakup beberapa teknik pemrosesan teks seperti normalisasi teks, case folding (mengubah semua teks menjadi huruf kecil), dan penghapusan stop words (kata-kata umum yang biasanya diabaikan dalam analisis teks). Proses ini membantu dalam menstandarkan teks ulasan sehingga lebih mudah untuk dianalisis.

4. Tokenisasi

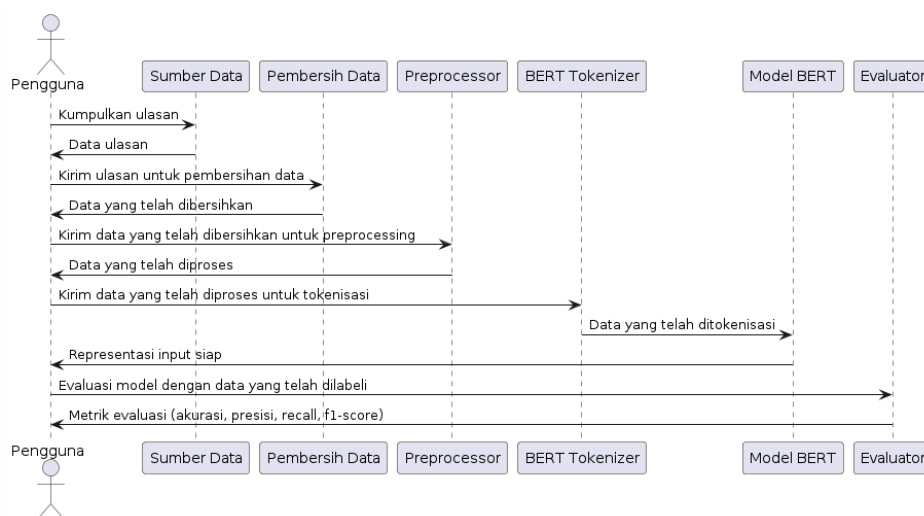
Data yang telah diproses kemudian dikirim ke BERT Tokenizer untuk tokenisasi. Pada kasus ini, digunakan BERTBASE-MULTILINGUAL-UNCASED Tokenizer. Tokenisasi adalah proses memecah teks ulasan menjadi token (yaitu, kata-kata atau sub-kata) yang dapat dimengerti oleh model BERT. Tokenisasi BERT mencakup penambahan token khusus seperti [CLS] di awal setiap kalimat dan [SEP] di akhir setiap kalimat, serta melakukan padding jika diperlukan. Hal ini membuat teks ulasan siap untuk dimasukkan ke dalam model BERT.

5. Penggunaan Model BERT

Data yang telah ditokenisasi kemudian dikirim ke model BERT. Pada kasus ini, digunakan model BERTBASE-MULTILINGUAL-UNCASED. Model ini telah dilatih sebelumnya untuk memahami berbagai bahasa tanpa membedakan huruf besar dan kecil, sehingga sangat cocok untuk platform X yang mungkin memiliki ulasan dalam berbagai bahasa. Model BERT mengubah data yang telah ditokenisasi menjadi representasi input yang dapat dipahami oleh model. Model BERT akan menganalisis aspek-aspek tertentu dalam ulasan (misalnya, aktung, plot) dan menentukan sentimen (positif, negatif, netral) dari setiap aspek tersebut.

6. Evaluasi Model

Setelah analisis sentimen dilakukan oleh model BERT, hasilnya dievaluasi menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan f1-score. Evaluasi ini dilakukan untuk mengukur kinerja model dalam mengidentifikasi aspek dan sentimen dari ulasan platform X. Hasil evaluasi memberikan informasi penting tentang seberapa baik model BERT dalam melakukan tugas analisis sentimen multi-aspek ini.



Gambar 5. Implementasi BERT

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data dikumpulkan dengan menggunakan teknik Crawl Data pada platform situs X. Crawl Data akan bekerja menggunakan Node.js. Oleh karena itu, peneliti perlu mengekstrak data tersebut ke dalam bentuk CSV, kemudian data dikumpulkan menjadi satu dokumen. Data yang akan dikumpulkan pada platform X berupa tweet dari beberapa user X oleh peneliti, setelah data dikumpulkan minimal sebanyak 3447 data tweet. Data akan disimpan dalam bentuk CSV dan siap untuk proses berikutnya.

```

○ ○ ○

# Crawl Data

filename = 'israel-palestine5.csv'
search_keyword = 'israel lang:id'
limit = 1000

!npx -y tweet-harvest@2.6.0 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token
{twitter_auth_token}

```

Gambar 6. Program Crawl Data

bagian-bagian dari perintah tersebut:

1. `npx tweet-harvest@2.6.0`: Ini adalah perintah yang dijalankan. `npx` adalah utilitas untuk menjalankan paket npm tanpa menginstalnya secara global. `tweet-harvest@2.6.0` adalah nama paket yang akan dijalankan, dengan versi spesifiknya.
2. `-o "israel-palestine5.csv"`: Ini adalah opsi perintah untuk menentukan nama file output dari hasil crawling data X. Dalam hal ini, file output akan disimpan dengan nama `'israel-palestine5.csv'`.
3. `-s "israel lang:id"`: Ini adalah opsi perintah untuk menentukan kata kunci pencarian. Data yang akan di-crawl adalah data yang terkait dengan kata kunci "israel" dalam bahasa Indonesia ('lang:id').
4. `--tab "LATEST"`: Ini adalah opsi perintah untuk menentukan tab atau bagian dari hasil pencarian X yang akan diambil. Dalam hal ini, kita hanya akan mengambil data dari tab "LATEST" atau yang terbaru.
5. `-l 1000`: Ini adalah opsi perintah untuk menentukan batas jumlah tweet yang akan diambil. Dalam contoh ini, kita akan mengambil maksimal 1000 tweet.
6. `--token {X_auth_token}`: Ini adalah opsi perintah untuk menentukan token otentikasi X yang diperlukan untuk mengakses API X. `{X_auth_token}` adalah placeholder untuk token otentikasi yang seharusnya diganti dengan token yang valid.

B. Cleansing Data

Berikut adalah kode Python sederhana untuk melakukan cleansing data teks setelah crawling. Kode ini menggunakan library ``re`` untuk ekspresi reguler dan ``nltk`` untuk pemrosesan bahasa alami:

```

import re
import nltk
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer

# Download stopwords dan wordnet dari nltk
nltk.download('stopwords')
nltk.download('wordnet')

# Inisialisasi objek lemmatizer dan stopwords
lemmatizer = WordNetLemmatizer()
stop_words = set(stopwords.words('english'))

def clean_text(text):
    # Hapus URL, mention, dan hashtag
    text = re.sub(r'http\S+|www\S+|https\S+|\@\w+|\#\w+', '', text)

    # Hapus tanda baca dan karakter khusus
    text = re.sub(r'[^\w\s]', '', text)

    # Ubah teks menjadi huruf kecil
    text = text.lower()

    # Hapus angka
    text = re.sub(r'\d+', '', text)

    # Hapus stopwords
    text = ' '.join([word for word in text.split() if word not in stop_words])

    # Lemmatization
    text = ' '.join([lemmatizer.lemmatize(word) for word in text.split()])

    return text

# Contoh penggunaan
sample_text = "This is an example tweet! Visit http://example.com @username #hashtag"
cleaned_text = clean_text(sample_text)
print(cleaned_text)

```

Gambar 7. Program Cleaning Data

Penjelasan langkah demi langkah:

1. Hapus URL, mention, dan hashtag: Menggunakan `re.sub` untuk menghapus pola URL (`http\S+`, `www\S+`, `https\S+`), mention (`@\w+`), dan hashtag (`#\w+`).
2. Hapus tanda baca dan karakter khusus: Menggunakan `re.sub` untuk menghapus karakter selain huruf dan spasi (`[^\w\s]`).
3. Ubah teks menjadi huruf kecil: Menggunakan `text.lower()` untuk mengonversi teks menjadi huruf kecil.
4. Hapus angka: Menggunakan `re.sub` untuk menghapus angka (`\d+`).
5. Hapus stopwords: Menggunakan `stopwords` dari `nltk` untuk menghapus kata-kata umum yang tidak signifikan.
6. Lemmatization: Menggunakan `WordNetLemmatizer` untuk mengubah kata ke bentuk dasarnya.

Setelah proses cleansing selesai proses akan dibagi menjadi 2 yaitu:

1. Pelabelan data yang dilakukan secara manual oleh peneliti
2. Tokenisasi yang akan di proses menggunakan BERT

C. Pelabelan Data

Labeling data dilakukan secara manual menggunakan excel. Tahap pelabelan data dilakukan secara mandiri oleh peneliti. Pelabelan dilakukan berdasarkan sentiment yang terkandung dalam tweet. Pemilihan sentimen dilakukan dengan melihat tweet yang di dapatkan dari proses Crawl Data. Untuk label sentimen di bagi menjadi 3 yaitu positive, netral, negative

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	ersation	created	avorite	cou	full_text	id_str	image_url	to_scre	lang	location	uote_cour	reply_count	tweet_cou	tweet_url	user_id	stusername	sentiment
2	1,79E+18	Tue May 1	0	Pengungs	1,79E+18	https://pbs.twimg.c	in				0	0	0	https://tw	3,86E+08	vthreeizm	negative
3	1,79E+18	Tue May 1	0	Pemukim	1,79E+18	https://pbs.twimg.c	in		I live in Te		0	0	0	https://tw	1,12E+18	Parstoday	negative
4	1,79E+18	Tue May 1	0	@_khanaz	1,79E+18		_khanazai	in			0	0	0	https://tw	3,03E+09	lucimaya1	negative
5	1,79E+18	Tue May 1	1	Jangan sei	1,79E+18			in	lost.at.sea		0	0	0	https://tw	3,66E+08	fallfortae	negative
6	1,79E+18	Tue May 1	0	@CNNInd	1,79E+18		CNNIndor	in	Random		0	0	0	https://tw	7,09E+17	peppew6	positive
7	1,79E+18	Tue May 1	0	Biadab! Is	1,79E+18			in			0	0	0	https://tw	1,23E+18	RepelitaO	negative
8	1,79E+18	Tue May 1	0	@convom	1,79E+18		convomfs	in	EXO PLAN		0	0	0	https://tw	7,32E+17	loeycrush	negative
9	1,79E+18	Tue May 1	0	gila ya ter	1,79E+18			in	fa+randon		0	1	0	https://tw	1,08E+18	dexkthv	neutral
10	1,79E+18	Tue May 1	0	sakit bgt s	1,79E+18			in	Doncaster		0	0	0	https://tw	2,43E+08	radiahami	negative

Gambar 8. Hasil Pelabelan Manual

D. Implementasi BERT

Implementasi BERT akan dilakukan beberapa tahapan, yakni tokenisasi, pemodelan, dan prediksi sentimen. Setelah dataset selesai melalui tahap cleansing data, kemudian model akan melakukan tokenisasi pada dataset. Proses tokenisasi menggunakan Tokenizer.

Penjelasan Langkah-Langkah

1. Inisialisasi Model dan Tokenizer:

Tokenisasi adalah proses mengubah teks mentah menjadi token yang dapat diproses oleh model pembelajaran mesin. Dalam konteks kode Anda, tokenisasi dilakukan oleh tokenizer dari model Transformer

```

○ ○ ○

model_name = "bert-base-uncased"
model = AutoModelForSequenceClassification.from_pretrained(model_name)
tokenizer = AutoTokenizer.from_pretrained(model_name)

```

Gambar 9. Program Tokenizer oleh BERT

- model_name adalah nama model yang digunakan.
- AutoModelForSequenceClassification dan AutoTokenizer menginisialisasi model dan tokenizer.

2. Fungsi sentiment_score:

```

○○○

def sentiment_score(data):
    tokens = tokenizer.encode(data, return_tensors='pt')
    print(f"Teks asli: {data}")
    print(f"Token: {tokens}")
    result = model(tokens)
    predicted_class = int(torch.argmax(result.logits))
    sentiment_labels = ["negative", "neutral", "positive"]
    if predicted_class == 0:
        return "negative"
    elif predicted_class == 1:
        return "neutral"
    else:
        return "positive"

```

Gambar 10. Program Tokenizer oleh BERT

- Mengambil teks input (data).
- Menggunakan `tokenizer.encode(data, return_tensors='pt')` untuk melakukan tokenisasi dan mengubah teks menjadi tensor PyTorch.
- Menjalankan tensor token melalui model (`result = model(tokens)`).
- Memprediksi kelas sentimen dengan melihat argmax dari logits yang dihasilkan (`predicted_class`).
- Mengembalikan label sentimen berdasarkan prediksi kelas.

3. Membaca Data dan Menerapkan Klasifikasi Sentimen:

```

○○○

df = pd.read_csv('/content/combined_data.csv')
print("Kolom yang tersedia dalam CSV:", df.columns)
column_name = 'full_text'
df['sentiment'] = df[column_name].apply(lambda x: sentiment_score(x))
df.to_csv('classified_texts.csv', index=False)

```

Gambar 11. Program Menerapkan Klasifikasi Sentimen Pada Data

- Membaca data dari file CSV.
- Menampilkan kolom yang tersedia dalam CSV.
- Menerapkan fungsi `sentiment_score` pada kolom `full_text` untuk mendapatkan sentimen dari setiap teks.
- Menyimpan hasil klasifikasi ke file CSV baru.

4. Menampilkan dan Menghitung Hasil Klasifikasi:


```

○○○

print("Hasil klasifikasi telah disimpan ke 'classified_texts.csv'.")
print(df[[column_name, "sentiment"]].head())
sentiment_counts = df['sentiment'].value_counts()
total_tweets = len(df)
sentiment_percentages = (sentiment_counts / total_tweets) * 100
print("Jumlah dan Persentase Sentimen:")
for sentiment, count in sentiment_counts.items():
    percentage = sentiment_percentages[sentiment]
    print(f"{sentiment.capitalize()}: {count} tweets ({percentage:.2f}%)")

```

Gambar 13. Proses editorial di Jurnal UMSIDA

- Menampilkan hasil klasifikasi dari beberapa baris pertama.
- Menghitung jumlah dan persentase dari setiap kelas sentimen.
- Menampilkan jumlah dan persentase dari setiap sentimen.

5. Contoh Output Tokenisasi

```

○○○

Teks asli: This is a sample tweet
Token: tensor([[ 101, 2023, 2003, 1037, 7099, 17942, 22942, 102]])

```

Gambar 12. Proses editorial di Jurnal UMSIDA

- 101, 2023, 2003, 1037, 7099, 17942, 22942, 102 adalah token numerik yang dihasilkan dari teks input oleh tokenizer.

E. Evaluasi

Dalam melakukan analisis sentimen menggunakan BERT, Model akan menentukan nilai akurasi setelah melalui pengujian kemudian akan mencoba untuk memprediksi hasil akurasi tertinggi secara keseluruhan menggunakan BERT. Hasil evaluasi yang disajikan mencakup beberapa informasi penting tentang kinerja model klasifikasi, seperti akurasi, precision, recall, dan F1-Score, serta confusion matrix yang memberikan gambaran lebih detail tentang performa model dalam mengklasifikasikan data.

Dari hasil evaluasi yang diberikan, kita dapat melihat beberapa informasi penting:

1. **Kolom di df_reference dan df_predicted: Menunjukkan kolom yang ada di kedua DataFrame yang akan digabungkan.**

```

○○○

Kolom di df_reference: Index(['Cleaned_Tweet', 'sentiment'], dtype='ob
Kolom di df_predicted: Index(['Cleaned_Tweet', 'sentiment'], dtype='ob

```

Gambar 14. Program Menampilkan Kolom

2. **Data gabungan: Merupakan contoh lima baris data setelah kedua**

DataFrame digabungkan. Dalam data gabungan ini, kita memiliki kolom `Cleaned_Tweet` yang berisi teks tweet, serta kolom `sentiment_reference` dan `sentiment_predicted` yang berisi label sentimen yang sesuai dengan teks tweet tersebut.

```

○ ○ ○

Data gabungan:

                                Cleaned_Tweet sentiment_reference
0  pemukim zionis mencegat dan menyerang truk yan...      negative
1  tidak seorang pun dari zionist yahudi laknat...      negative
2  jangan sembarangan ngeluarin opini kalo km pak...      negative
3  jangan pernah lelah bicara tentang tanah pale...      positive
4  biadab israel sengaja bunuh para ulama untuk l...      negative

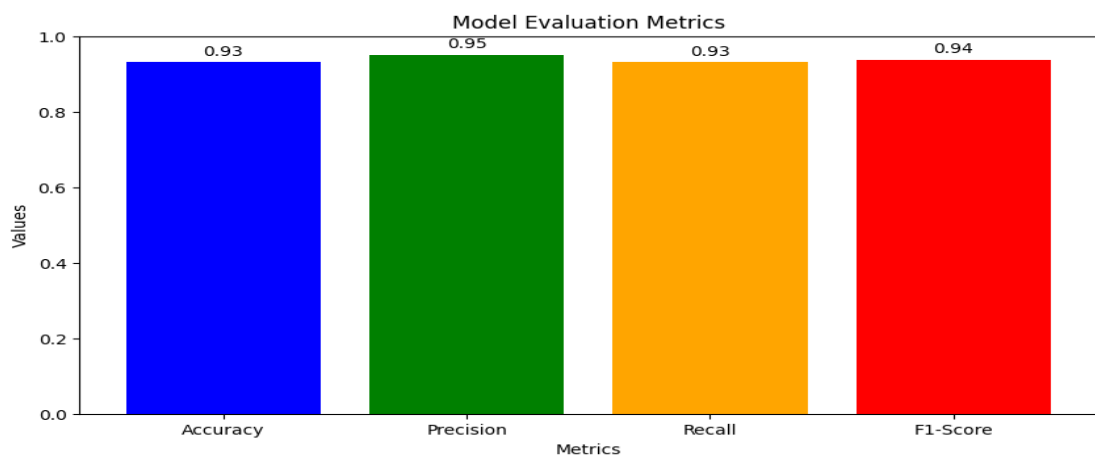
sentiment_predicted
0      negative
1      negative
2      negative
3      positive
4      negative

```

Gambar 15. Hasil Sentimen

3. Hasil Evaluasi Model: Menunjukkan hasil evaluasi performa model

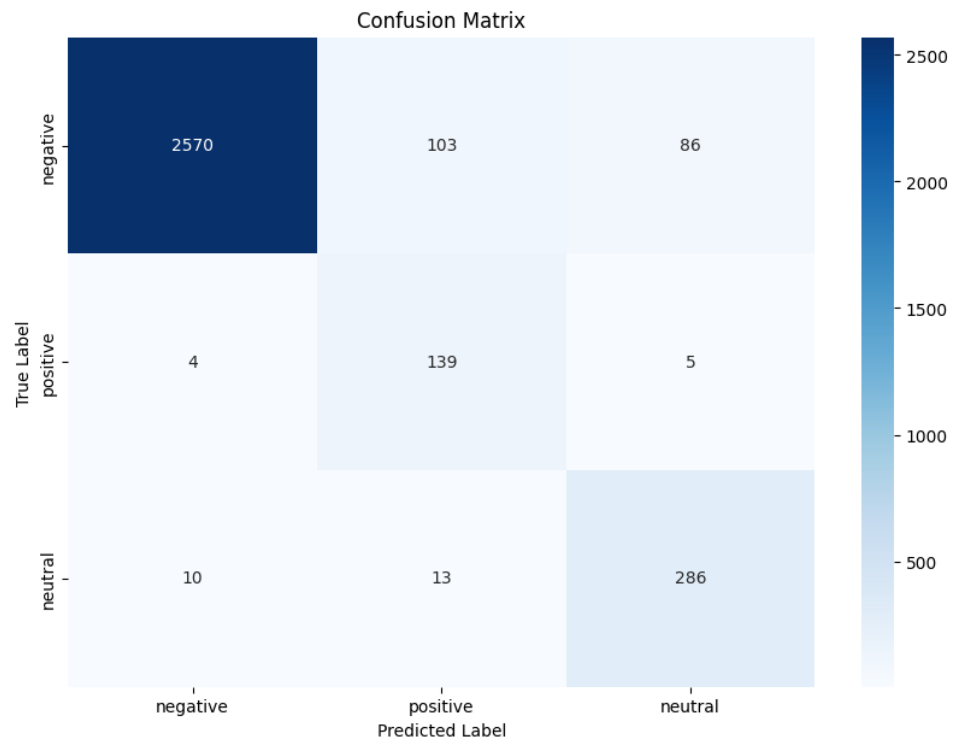
klasifikasi sentimen. Di sini, kita memiliki nilai akurasi sebesar 0.93, presisi (rata-rata tertimbang) sebesar 0.95, recall (rata-rata tertimbang) sebesar 0.93, dan F1-Score (rata-rata tertimbang) sebesar 0.94. Semakin tinggi nilai akurasi, presisi, recall, dan F1-Score, semakin baik performa model klasifikasi sentimen tersebut.



Gambar 16. Hasil Evaluasi

4. Confusion Matrix: Merupakan matriks yang menunjukkan jumlah prediksi

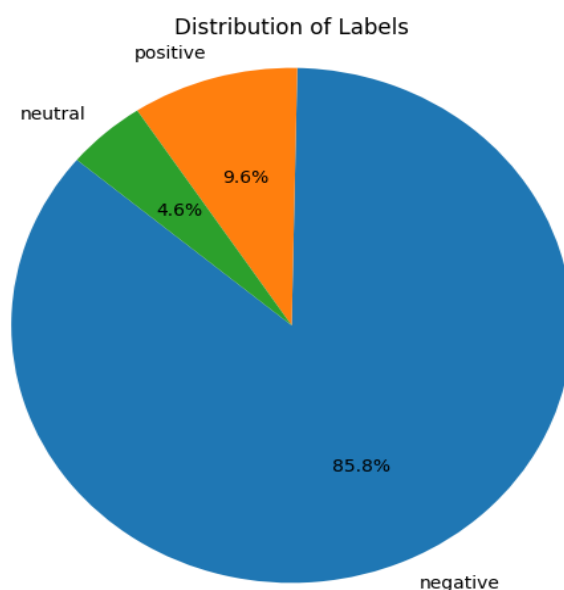
yang benar dan salah untuk setiap kelas sentimen. Baris pada confusion matrix menunjukkan true labels (label sebenarnya), sedangkan kolomnya menunjukkan predicted labels (label yang diprediksi oleh model). Dari confusion matrix, kita dapat melihat bahwa:



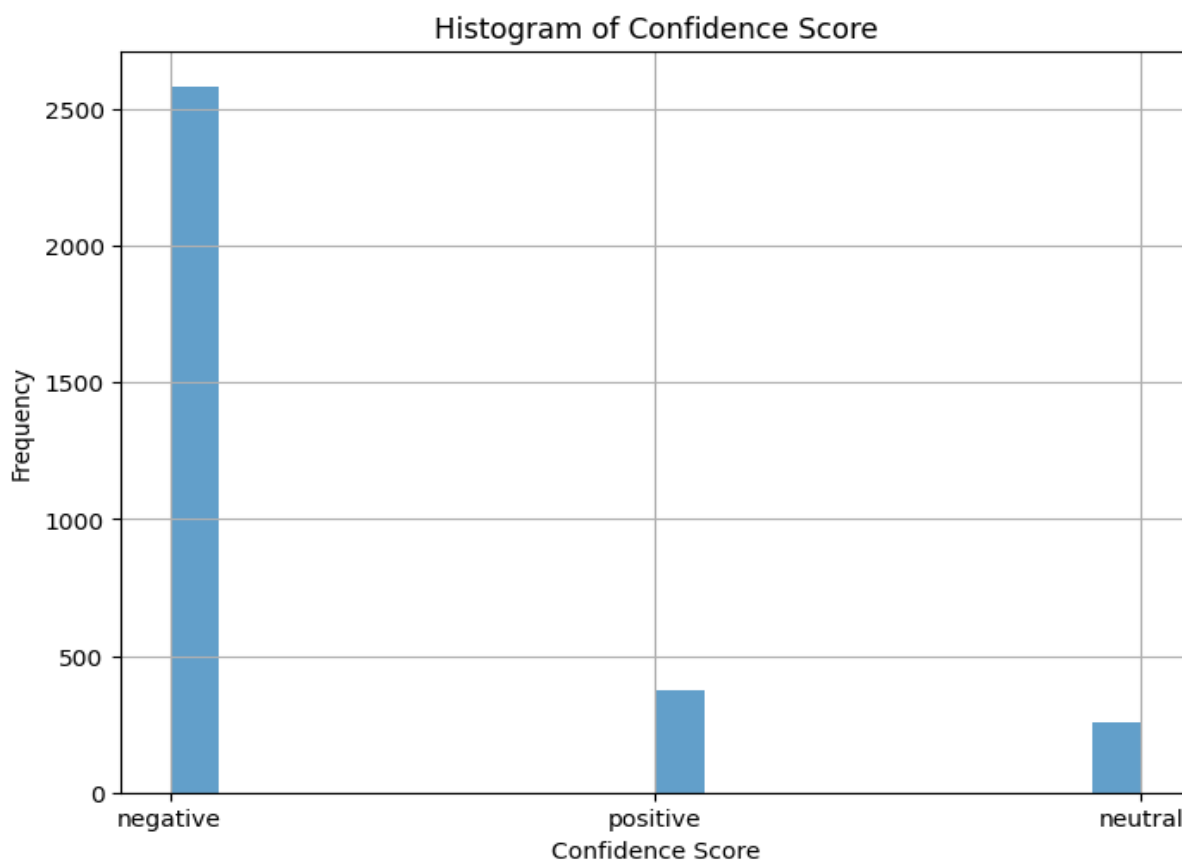
Gambar 17. Tabel Confusion Matrix

- Untuk kelas pertama (sentimen negatif), terdapat 2570 prediksi yang benar, 103 prediksi yang salah sebagai sentimen netral, dan 86 prediksi yang salah sebagai sentimen positif.
- Untuk kelas kedua (sentimen netral), terdapat 139 prediksi yang benar, 4 prediksi yang salah sebagai sentimen negatif, dan 5 prediksi yang salah sebagai sentimen positif.
- Untuk kelas ketiga (sentimen positif), terdapat 286 prediksi yang benar, 10 prediksi yang salah sebagai sentimen negatif, dan 13 prediksi yang salah sebagai sentimen netral.

Dengan demikian, hasil evaluasi ini memberikan gambaran tentang seberapa baik model klasifikasi sentimen dalam memprediksi label sentimen pada data tweet yang diberikan.



Gambar 18. Tabel Distribution



Gambar 19. Tabel Histogram

II. KESIMPULAN

Penerapan sentimen analisis berfokus `nlptown/bert-base-multilingual-uncased-sentiment` untuk melihat seberapa akurat BERT model terbaru dalam menentukan penilaian terhadap ulasan berupa Tweet. Penelitian ini melakukan beberapa percobaan dengan menggunakan metode NLP (Natural Language Processing). Hasil evaluasi menggunakan metode NLP menunjukkan bahwa model klasifikasi sentimen telah memberikan kinerja yang memuaskan dalam memprediksi sentimen dari teks. Dari akurasi sebesar 0.93, kita dapat menyimpulkan bahwa sekitar 93% dari total prediksi yang dilakukan oleh model adalah benar. Namun, penting untuk memahami performa model secara lebih rinci dengan melihat presisi, recall, dan F1-Score. Presisi model, yang mengukur seberapa banyak dari prediksi positif yang sebenarnya benar, mencapai nilai 0.95. Ini berarti sebagian besar dari teks yang diprediksi sebagai positif oleh model memang memiliki sentimen positif. Namun demikian, nilai recall sebesar 0.93 menunjukkan bahwa model mampu mengidentifikasi sebagian besar teks dengan sentimen positif yang sebenarnya. F1-Score, yang merupakan nilai rata-rata harmonik antara presisi dan recall, mencapai 0.94, menunjukkan keseimbangan yang baik antara presisi dan recall model. Namun, meskipun model telah memberikan hasil yang baik secara keseluruhan, kita juga perlu memeriksa confusion matrix untuk melihat detail performa model untuk setiap kelas sentimen. Dari confusion matrix, kita dapat melihat bahwa sebagian besar prediksi adalah benar, terutama untuk kelas sentimen negatif dan positif. Namun, masih terdapat beberapa prediksi yang salah, terutama pada kelas sentimen netral. Oleh karena itu, walaupun model telah memberikan hasil yang memuaskan, masih ada ruang untuk peningkatan melalui pengembangan teknik NLP yang lebih canggih dan penanganan data yang lebih baik. Dengan demikian, model ini dapat digunakan dengan percaya diri untuk menganalisis sentimen pengguna dalam berbagai konteks, seperti analisis media sosial, pemantauan merek, dan pengelolaan reputasi perusahaan, namun perbaikan lebih lanjut tetap diperlukan untuk memaksimalkan keakuratan dan kinerja model di masa mendatang.

UCAPAN TERIMA KASIH

Penelitian ini sekarang telah berakhir, dan saya ingin mengucapkan terima kasih kepada semua pihak yang telah berperan serta dalam perjalanan ini. Terima kasih kepada keluarga saya atas doa, dukungan, dan kesabaran yang mereka berikan selama proses ini. Terima kasih juga kepada teman-teman dan rekan-rekan sejawat atas dukungan moral dan semangat yang selalu mereka berikan. Dan khususnya, terima kasih kepada dosen pembimbing saya, atas bimbingan, arahan, dan kesabaran yang luar biasa selama penelitian ini. Dukungan dari beliau telah membantu saya melewati setiap tahap penelitian ini. Tak lupa, terima kasih kepada Kaprodi dan seluruh staf di Program Studi Teknik Informatika di Universitas Muhammadiyah Sidoarjo (UMSIDA), atas dukungan dan bimbingannya. Semoga hasil penelitian ini dapat memberikan kontribusi yang berarti bagi kemajuan ilmu pengetahuan. Terima kasih atas semua dukungan yang telah diberikan.

REFERENSI

- [1] Akhmad, E. P. A. (2023). Analisis Sentimen Ulasan Aplikasi DLU Ferry Pada Google Play Store Menggunakan Bidirectional Encoder Representations from Transformers. *JURNAL APLIKASI PELAYARAN DAN KEPELABUHANAN*, 13(2), 104–112. <https://doi.org/10.30649/japk.v13i2.94>
- [2] Chandradev, V., Made, I., Dwi Suarjaya, A., Putu, I., & Bayupati, A. (n.d.). Chandradev, Analisis Sentimen Review Hotel menggunakan Metode Deep Learning BERT 107 Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT.
- [3] Karimah, N., & Baita, A. (2024). Multi-Aspect Sentiment Analysis Pada Review Film Menggunakan Metode Bidirectional Encoder Representations From Transformers (BERT) Multi-Aspect Sentiment Analysis of Film Review Using Bidirectional Encoder Representations from Transformers (BERT). *Jurnal Sistem Komputer*, 13(1), 2020. <https://doi.org/10.34010/komputika.v13i1.11098>
- [4] Khatib Sulaiman, J., Putri, M., Edy Sutanto, T., & Inna, S. (n.d.). Studi Empiris Model BERT dan DistilBERT: Analisis Sentimen pada Pemilihan Presiden Indonesia. *Indonesian Journal of Computer Science Attribution*, 12(5), 2023–2972
- [5] Putu, N., Saraswati, V. D., Yudistira, N., & Adikara, P. P. (2023). Analisis Sentimen terhadap Perundangan Siber pada X menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT) (Vol. 7, Issue 2). <http://j-ptiik.ub.ac.id>.
- [6] Rifqah, *, Tanjung, R., Satria, A., Rahmayani Tanjung, R., Sirait, R. I., & Sihotang, J. J. (2024). Konflik Palestina: Jihad Netizen Indonesia, Solidaritas Atau Pelanggaran Hukum. 2(2), 229–238. <https://doi.org/10.51903/jaksa.v1i3.14670>
- [7] Sofi, N., Sulistyorini, T., & Nazaruddin, M. (n.d.). Analisis Sentimen Masyarakat Pengguna Media Sosial X Terhadap Motogp Mandalika Lombok Menggunakan Metode Bidirectional Encoder Representation From Transformers (BERT). 1, 2023.
- [8] Sagita Dewi Manajemen Bisnis Syari, R., & Hamfara Yogyakarta, S. (2024). PENGARUH KONFLIK PALESTINA-ISRAEL TERHADAP PEREKONOMIAN DUNIA (Vol. 2).
- [9] Sofi, N., Sulistyorini, T., & Nazaruddin, M. (n.d.). Analisis Sentimen Masyarakat Pengguna Media Sosial Twitter Terhadap Motogp Mandalika Lombok Menggunakan Metode Bidirectional Encoder Representation From Transformers (BERT). 1, 2023.
- [10] Tri Wicaksono, A., Arbi, A., Badrotin Jabbar, N., Fajruddin Fatwa, A., Ahmad Yani, J., Wonocolo No, K., & Korespondensi penulis, S. (2024). Problematika ICC Dalam Menjatuhkan Sanksi Kepada Israel Dalam Perspektif Hukum Internasional. *JHPIS*, 3(1), 207–224. <https://doi.org/10.55606/jhpis.v3i1.3210>
- [11] Trisnawati Manajemen Bisnis Syariah, R., & Hamfara, S. (n.d.). BOIKOT DAN AKTIVISME: PERILAKU KONSUMEN DALAM ISU KONFLIK ISRAEL-PALESTINA.
- [12] Muktafin, E. H., Kusri, K., & Luthfi, E. T. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Jurnal Eksplora Informatika*, 10(1), 32-42
- [13] Fathoniah, S., & Rozikin, C. (2022). Analisis Sentimen Masyarakat terhadap Teroris dalam Media Sosial Twitter menggunakan NLP. *Jurnal Ilmiah Wahana Pendidikan*, 8(13), 412-419
- [14] Munasatya, N., & Novianto, S. (2020). Natural Language Processing untuk Analisis Sentimen Presiden Jokowi Menggunakan Multi Layer Perceptron. *Techno. Com*, 19(3).
- [15] Sumendapa, R. V., & Mahendra, I. B. M. (2023). Membandingkan Analisis Sentimen Review Pelanggan Shopee Dan Tokopedia Menggunakan Google's NLP API. *Jurnal Elektronik Ilmu Komputer Udayana p-ISSN*, 2301, 5373

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.