ANALISIS SENTIMEN OPINI PUBLIK TERHADAP BERITA PERANG IRAN-ISRAEL DI MEDIA SOSIAL PADA TAHUN 2026 MENGGUNAKAN MODEL INDOBERT

Maria Ulfa, Ridwan Yusuf
Teknik Informatika, Universitas Dharma Wacana Metro
Jl. Kenanga No.3, Mulyojati, Kec. Metro Barat, Kota Metro, Lampung, Indonesia
Email: 22010064@students.dharmawacana.ac.id

ABSTRAK

Konflik antara Iran dan Israel yang banyak diberitakan di berbagai platform digital memunculkan beragam opini publik yang tersebar melalui media sosial seperti X dan TikTok. Interaksi pengguna dalam bentuk komentar, diskusi, dan reaksi emosional menghasilkan data opini dalam jumlah besar yang bersifat tidak terstruktur dan dinamis. Kondisi ini membuat analisis manual menjadi tidak efisien dan tidak efektif. Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap konflik Iran–Israel menggunakan pendekatan *Natural Language Processing* (NLP). Metode yang digunakan adalah analisis sentimen dengan model IndoBERT untuk mengetahui sentimen masyarakat terhadap berita yang sedang populer di X dan TikTok. Data penelitian berupa komentar, balasan, dan unggahan pengguna media sosial yang berkaitan dengan konflik Iran–Israel dalam jangka waktu tertentu. Hasil klasifikasi menunjukkan kinerja yang sangat baik dari model IndoBERT dengan tingkat akurasi sebesar 96,86%. Hal ini menunjukkan bahwa model IndoBERT mampu memahami konteks bahasa Indonesia, termasuk teks informal dan tidak terstruktur yang sering ditemukan di media sosial.

Kata kunci : *Analisis Sentimen Opini Publik, Berita Perang Iran-Israel, Media Sosial, Model IndoBERT*

1. PENDAHULUAN

Dengan kemajuan teknologi informasi saat ini, masyarakat dapat dengan cepat mengakses berita global, termasuk konflik perang. Salah satu konflik yang menarik perhatian dunia adalah konflik antara Iran dan Israel yang menimbulkan ketegangan di Timur Tengah serta kekhawatiran terhadap stabilitas keamanan internasional [1]. Tingginya intensitas pemberitaan menjadikan konflik tersebut sebagai salah satu topik yang banyak diperbincangkan masyarakat dan menjadi trending di berbagai platform digital. Media sosial seperti X dan TikTok turut berperan dalam mempercepat penyebaran informasi serta menjadi ruang bagi masyarakat untuk menyampaikan opini dalam bentuk komentar, diskusi, maupun reaksi emosional terhadap berita yang sedang berkembang.

Banyaknya interaksi pengguna di media sosial menghasilkan data opini dalam jumlah besar yang bersifat tidak terstruktur dan dinamis. Karakteristik data tersebut menyebabkan proses analisis secara manual menjadi sulit, tidak efisien, dan membutuhkan waktu yang lama. Selain itu, penggunaan bahasa informal pada media sosial juga menjadi tantangan dalam proses analisis sentimen. Oleh karena itu, diperlukan pendekatan berbasis *Natural Language Processing* (NLP) untuk mengolah dan menganalisis opini publik secara otomatis dan sistematis [2].

Beberapa penelitian sebelumnya telah menggunakan metode seperti Naïve Bayes, Support Vector Machine (SVM), dan Random Forest untuk melakukan analisis sentimen pada media sosial [3], [2]. Namun, metode tersebut masih memiliki keterbatasan dalam memahami konteks bahasa secara menyeluruh, terutama pada data media sosial yang bersifat informal dan dinamis. Seiring perkembangan teknologi, model berbasis *deep learning* seperti IndoBERT mulai digunakan karena mampu memahami konteks bahasa Indonesia dengan lebih baik melalui pendekatan *transformer* [4]. Penelitian sebelumnya juga menunjukkan bahwa IndoBERT memiliki tingkat akurasi yang lebih baik dibandingkan metode konvensional dalam analisis sentimen media sosial [5], [6].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap konflik Iran–Israel pada media sosial X dan TikTok menggunakan model IndoBERT. Penelitian ini menggunakan pendekatan *Natural Language Processing* (NLP) dan metode analisis sentimen untuk mengklasifikasikan opini publik ke dalam kategori sentimen tertentu, sehingga dapat memberikan gambaran mengenai tanggapan masyarakat terhadap konflik Iran–Israel di media sosial.

2. TINJAUAN PUSTAKA

2.1. Media Sosial

Media sosial adalah platform digital berbasis internet yang memungkinkan orang berinteraksi, berbagi informasi, dan menemukan tanggapan terhadap masalah dalam waktu nyata. Perkembangan media sosial dalam beberapa tahun terakhir telah mengubah pola komunikasi masyarakat menjadi lebih terbuka dan partisipatif. Informasi dapat tersebar dengan sangat cepat melalui unggahan, komentar, maupun fitur berbagi, menjadikan media sosial sebagai tempat untuk diskusi publik yang aktif dan dinamis [3].

2.2. Opini Publik

Opini publik merupakan pandangan, sikap, atau penilaian masyarakat terhadap suatu isu, peristiwa,

atau informasi tertentu yang berkembang di ruang publik. Dalam konteks media sosial, opini publik umumnya disampaikan dalam bentuk teks singkat, komentar, maupun unggahan yang bersifat spontan. Karakteristik tersebut mencerminkan respons masyarakat yang cepat dan dinamis terhadap suatu berita atau isu yang sedang menjadi perhatian [7].

2.3. Analisis Sentimen

Analisis sentimen adalah komponen opinion mining yang secara khusus berkonsentrasi pada mengkategorikan polaritas opini ke dalam kategori tertentu, seperti positif, negatif, atau netral. Ini berbeda dengan cara opinion mining berusaha menggali opini secara keseluruhan, sehingga analisis sentimen lebih fokus pada menemukan kecenderungan sikap yang terkandung dalam teks dalam penelitian ini [8].

2.4. *Opinion Mining*

Buku Analisis Sentimen dan Penggalian Opini mengatakan bahwa penggalian opini tidak hanya menemukan polaritas sentimen; itu juga mencari elemen yang dibahas, orang yang dibicarakan, dan hubungan antara opini dan objek tersebut [5]. Dengan kata lain, penggalian pendapat memiliki lingkup yang lebih luas daripada analisis sentimen karena melibatkan proses penggalian informasi subjektif secara menyeluruh.

Karena opini masyarakat tersebar luas dan tidak terstruktur, opinion mining menjadi penting di media sosial. Oleh karena itu, metode komputasional diperlukan untuk memahami dan mengekstraksi opini yang berkembang secara sistematis.

2.5. *Natural Language Processing (NLP)*

Cabang kecerdasan buatan yang dikenal sebagai pengolahan bahasa alami (NLP) berfokus pada bagaimana bahasa alami dapat dipahami dan diproses oleh sistem komputer. NLP menjadi dasar dalam analisis sentimen karena seluruh proses pengolahan data teks dilakukan melalui tahapan-tahapan NLP.

Proses NLP mencakup beberapa tahapan, seperti tokenisasi, normalisasi teks, penghapusan kata tidak relevan, serta pembentukan representasi teks dalam bentuk numerik sehingga model klasifikasi dapat memprosesnya [2]. Tahapan ini sangat penting karena memungkinkan olahan sistematis data teks yang awalnya tidak terstruktur. Dalam penelitian ini, NLP berperan sebagai tahap pengolahan awal sebelum dilakukan klasifikasi sentimen menggunakan model IndoBERT.

2.6. Model *IndoBERT*

Metode baru untuk menganalisis sentimen berbasis teks ditawarkan oleh kemajuan model bahasa berbasis transformer. IndoBERT adalah model bahasa pra-latih khusus untuk bahasa Indonesia yang menggunakan arsitektur BERT (Bidirectional Encoder Representations from Transformers). Dengan kemampuan model ini untuk memahami konteks kata secara dua arah (bidirectional), representasi makna kata tidak lagi bergantung pada satu kata, melainkan bergantung pada keseluruhan struktur kalimat. Pendekatan berbasis contextual embedding ini memungkinkan sistem memahami makna kata sesuai konteks penggunaannya dalam kalimat [4].

Kemampuan memahami konteks secara mendalam ini menjadi alasan utama pemilihan IndoBERT dalam penelitian ini. Berita trending di media sosial umumnya memunculkan respons yang beragam, cepat, dan dinamis. Oleh karena itu, diperlukan model yang mampu memahami relasi semantik dan konteks kalimat secara menyeluruh. Dengan karakteristik tersebut, IndoBERT dinilai lebih sesuai dibandingkan metode klasifikasi tradisional untuk menganalisis sentimen opini publik terhadap berita trending di media sosial [11].

2.7. Pengumpulan Data Media Sosial

Karena tahapan awal yang sangat penting dalam penelitian analisis sentimen adalah pengumpulan data, hasil klasifikasi akan dipengaruhi oleh kualitas data yang diperoleh. Dalam penelitian ini, data opini publik dikumpulkan dari media sosial, yang berfungsi sebagai sumber utama percakapan tentang topik berita yang sedang dibahas. Metode seperti web scraping atau penggunaan Application Programming Interface (API) memungkinkan proses pengambilan data yang lebih otomatis. Metode ini memungkinkan pengumpulan data yang lebih efektif dan sistematis [9].

Dengan demikian, pengumpulan data dan prapemrosesan menjadi fondasi utama dalam penelitian ini sebelum dilakukan tahap pemodelan menggunakan IndoBERT.

2.8. Penelitian Terdahulu

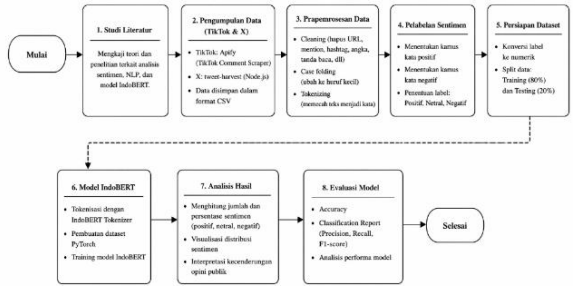
Beberapa penelitian sebelumnya telah dilakukan terkait analisis sentimen pada media sosial menggunakan berbagai metode klasifikasi. Penelitian oleh D. I. Putri menunjukkan bahwa model IndoBERT memiliki performa yang lebih baik dibandingkan metode konvensional dalam menganalisis sentimen pada media sosial X [9]. Selain itu, penelitian oleh F. Sugandi juga menunjukkan bahwa model berbasis BERT mampu menghasilkan klasifikasi sentimen dengan akurasi yang baik [10].

Penelitian lain oleh Siti Rihastuti dan Afnan Rosyidi menggunakan metode Random Forest untuk menganalisis sentimen pengguna TikTok terhadap progres pembangunan IKN [11]. Sementara itu, penelitian mengenai analisis sentimen multi-kelas menunjukkan bahwa klasifikasi sentimen dapat dilakukan ke dalam beberapa kategori untuk menghasilkan analisis yang lebih rinci [12].

Berdasarkan penelitian terdahulu tersebut, penelitian ini menggunakan model IndoBERT untuk menganalisis sentimen opini publik terhadap konflik Iran–Israel pada media sosial X dan TikTok.

3. METODE PENELITIAN

Metodologi penelitian ini menguraikan jenis dan pendekatan penelitian, objek atau subjek penelitian, teknik pengumpulan data, teknik analisis data, serta alur atau tahapan penelitian yang digunakan dalam menganalisis sentimen opini publik terhadap berita yang sedang trending di media sosial berbasis model IndoBERT.



Gambar 1. Alur Penelitian

3.1. Jenis dan Pendekatan Penelitian

Penelitian analistis ini menggunakan metode pemrosesan bahasa alami (Natural Language Processing). Penelitian ini bersifat deskriptif-analitis, karena bertujuan untuk menggambarkan dan menganalisis kecenderungan sentimen opini publik terhadap berita trending di media sosial.

3.2. Objek atau Subjek Penelitian

Fokus penelitian ini adalah sentimen opini publik terkait dengan sebuah berita yang sedang trending di media sosial. Subjek penelitian berupa data teks opini pengguna media sosial, yang berasal dari komentar, balasan, atau unggahan pengguna yang berkaitan dengan berita trending pada platform X dan TikTok dalam periode waktu tertentu.

3.3. Teknik Pengumpulan Data

Pada penelitian ini, proses pengumpulan data diawali dengan studi literatur melalui penelaahan jurnal, buku, dan penelitian terdahulu yang berkaitan dengan analisis sentimen, Natural Language Processing (NLP), serta model IndoBERT untuk memahami konsep dan metode yang digunakan. Selanjutnya, peneliti mengumpulkan data opini publik dari media sosial, khususnya TikTok dan X, terkait berita konflik Iran–Israel yang sedang trending. Pengambilan data dari TikTok dilakukan menggunakan layanan berbasis web Apify dengan memanfaatkan tools TikTok Comment Scraper, di mana peneliti terlebih dahulu mencari konten relevan, menyalin tautannya, lalu menjalankan proses crawling untuk mengumpulkan data komentar secara otomatis.

Selain menggunakan platform TikTok, penelitian ini juga menggunakan data dari media sosial X sebagai sumber opini publik. Proses pengambilan data dilakukan secara otomatis menggunakan tools berbasis Node.js, yaitu *tweet-harvest*. Pengumpulan data dilakukan dengan menentukan kata kunci yang relevan dengan topik penelitian, yaitu konflik Iran–Israel, serta membatasi jumlah data yang diambil untuk menjaga efisiensi proses crawling. Data yang diperoleh kemudian disimpan dalam format CSV dan digunakan sebagai dataset awal dalam proses analisis sentimen.

Implementasi proses crawling ditunjukkan pada Kode 1:

```
twitter_auth_token = 'YOUR_TOKEN'

filename = 'data_twitter.csv'
search_keyword = 'Iran Israel'
limit = 900

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}

import pandas as pd
df = pd.read_csv(f"tweets-data/{filename}")

print("Jumlah data:", len(df))
```

3.4. Teknik Analisis Data

Tahapan analisis data yang dilakukan dalam penelitian ini meliputi:

3.5. Prapemrosesan Data Teks

Pada tahap ini, data komentar dari X dan TikTok masih berupa data mentah yang mengandung elemen tidak relevan seperti URL, mention, hashtag, angka, dan tanda baca berlebih, sehingga belum siap dianalisis. Oleh karena itu, dilakukan prapemrosesan menggunakan bahasa pemrograman Python melalui Google Colab, yang meliputi pembersihan teks (cleaning) untuk menghilangkan elemen yang tidak diperlukan, normalisasi teks (case folding) dengan mengubah seluruh huruf menjadi kecil, serta tokenizing untuk memecah teks menjadi kumpulan kata. Implementasi proses prapemrosesan ditunjukkan pada Kode 2:

```
        def clean_text(text):
            text = text.lower()
            text = re.sub(r'http\S+', '', text)
            text = re.sub(r'@\w+', '', text)
            text = re.sub(r'#\w+', '', text)
            text = re.sub(r'\d+', '', text)
            text = text.translate(str.maketrans('', '', string.punctuation))
            return text

        def tokenize(text):
            return text.split()

        df['clean_text']= df['text'].apply(clean_text)
    df['tokens'] = df['clean_text'].apply(tokenize)
```

Hasilnya, data yang semula tidak terstruktur berhasil diubah menjadi lebih bersih, rapi, dan siap digunakan untuk analisis sentimen pada tahap selanjutnya.

3.6. Pelabelan Data Sentimen

Setelah melalui tahap prapemrosesan, seluruh data komentar diberi label sentimen secara otomatis menggunakan pendekatan berbasis leksikon dengan bantuan program Python. Peneliti terlebih dahulu menyusun daftar kata yang merepresentasikan sentimen positif dan negatif, kemudian setiap komentar dianalisis berdasarkan jumlah kemunculan kata-kata tersebut. Komentar dikategorikan sebagai sentimen positif apabila kata positif lebih dominan, sebagai sentimen negatif apabila kata negatif lebih dominan, dan sebagai sentimen netral apabila tidak terdapat dominasi di antara keduanya.

Implementasi proses pelabelan sentimen ditunjukkan pada Kode 3:

```
# daftar kata sentimen
positive_words = [
   "baik", "bagus", "hebat", "kuat", "damai",
   "aman", "bangga", "tenang", "senang"
]

negative_words = [
   "buruk", "parah", "takut", "serangan",
   "perang", "hancur", "ancaman", "bahaya",
   "tegang", "konflik"
]

def label_sentiment(text):
   text = text.lower()

   pos = sum(word in text for word in
positive_words)
   neg = sum(word in text for word in
negative_words)

   if pos > neg:
      return "positif"
   elif neg > pos:
      return "negatif"
   else:
      return "netral"

# menerapkan ke dataset
df['sentiment'] =
df['clean_text'].apply(label_sentiment)
```

3.7. Penerapan Model IndoBERT

Pada tahap ini, peneliti melakukan proses klasifikasi sentimen menggunakan model IndoBERT. Sebelum proses pemodelan dilakukan, peneliti terlebih dahulu menyiapkan dataset yang akan digunakan.

Dataset dimuat dari file berformat CSV menggunakan library *pandas*. Setelah data berhasil dimuat, dilakukan pengecekan awal dengan menampilkan beberapa baris pertama dari dataset untuk memastikan bahwa data telah terbaca dengan baik dan siap digunakan pada tahap selanjutnya. Implementasi proses pemuatan dataset ditunjukkan pada Kode 4:

```
import pandas as pd
import torch

from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,
classification_report

from transformers import AutoTokenizer,
AutoModelForSequenceClassification, Trainer,
TrainingArguments

from google.colab import files

uploaded = files.upload()

# membaca dataset
df = pd.read_csv('HASIL STRAT.csv')
print(df.head())
```

Selanjutnya, label sentimen yang semula berbentuk teks seperti “negatif”, “netral”, dan “positif” diubah menjadi bentuk numerik agar dapat diproses oleh model. Proses ini dilakukan karena model klasifikasi tidak dapat mengolah label dalam bentuk teks secara langsung. Oleh karena itu, label negatif direpresentasikan sebagai 0, netral sebagai 1, dan positif sebagai 2. Implementasi proses konversi label ditunjukkan pada Kode 5:

```
# konversi label ke angka
mapping = {
   'negatif': 0,
   'netral': 1,
   'positif': 2
}

df['sentimen_angka'] = df['sentiment'].map(mapping)

texts = df['clean_text'].astype(str)
labels = df['sentimen_angka']
```

Setelah itu, peneliti mengambil dua bagian penting dari dataset, yaitu teks yang sudah dibersihkan (clean_text) sebagai data utama, dan label sentimen sebagai target. Kemudian data tersebut dibagi menjadi data latih dan data uji dengan perbandingan 80% untuk pelatihan dan 20% untuk pengujian. Pembagian data ini bertujuan untuk melatih model serta mengevaluasi kinerjanya pada data yang belum pernah dilihat sebelumnya. Implementasi proses pembagian data ditunjukkan pada kode 6

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    texts,
    labels,
    test_size=0.2,
    random_state=42
)
```

Setelah proses pembagian data, langkah selanjutnya adalah memanggil tokenizer IndoBERT yang digunakan untuk mengubah teks menjadi bentuk yang dapat dipahami oleh model. Implementasi pemanggilan tokenizer ditunjukkan pada Kode 7:

```
from transformers import BertTokenizer

tokenizer =
BertTokenizer.from_pretrained('indobenchmark/indo
bert-base-p1')
```

Langkah berikutnya, peneliti memanggil model IndoBERT yang sudah tersedia. Model ini sebelumnya sudah dilatih, sehingga pada penelitian ini peneliti hanya menyesuaikan model dengan jumlah kelas sentimen yang digunakan. Implementasi proses pelatihan model ditunjukkan pada Kode 8:

```
from transformers import
BertForSequenceClassification

model=BertForSequenceClassification.from_pretrain
ed(
    'indobenchmark/indobert-base-p1',
    num_labels=3
)
```

Terakhir, peneliti menjalankan proses pelatihan model. Pada tahap ini, model dilatih menggunakan data latih untuk mengenali pola kalimat yang termasuk ke dalam kategori sentimen positif, negatif, atau netral. Proses pelatihan dilakukan dalam beberapa iterasi (*epoch*) agar model dapat mempelajari pola data dengan lebih baik. Implementasi proses pelatihan model ditunjukkan pada Kode 9:

```
from transformers import Trainer,
TrainingArguments

training_args = TrainingArguments(
    output_dir='./results',
    num_train_epochs=2,
    per_device_train_batch_size=8,
    per_device_eval_batch_size=8,
    logging_dir='./logs'
)

trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=train_dataset,
    eval_dataset=test_dataset
)

trainer.train()
```

3.8. Analisis Hasil Klasifikasi

Setelah proses pelatihan selesai, model diuji menggunakan data uji yang telah dipisahkan untuk memprediksi sentimen setiap komentar ke dalam kategori positif, negatif, atau netral. Hasil prediksi tersebut kemudian dibandingkan dengan data asli guna mengevaluasi tingkat ketepatan model.

Berdasarkan hasil ini, peneliti menganalisis kecenderungan sentimen yang muncul, seperti dominasi opini negatif, positif, atau netral, sehingga dapat diketahui respons masyarakat terhadap isu yang dibahas. Selain itu, peneliti juga mengidentifikasi pola umum dalam komentar guna memperoleh gambaran menyeluruh mengenai opini publik terhadap konflik Iran–Israel di media sosial.

3.9. Evaluasi Model

Pada tahap terakhir, peneliti mengecek seberapa bagus model yang sudah dibuat. Caranya dengan membandingkan hasil prediksi model dengan data sebenarnya.

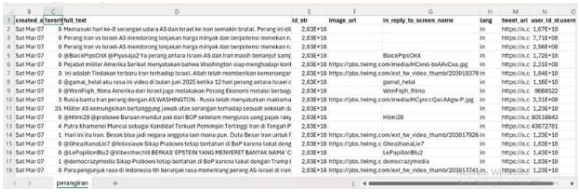
Peneliti menghitung nilai accuracy untuk melihat berapa banyak prediksi yang benar dari keseluruhan data. Semakin tinggi nilainya, berarti model semakin baik. Implementasi proses evaluasi model ditunjukkan pada Kode 10.

```
    predictions = trainer.predict(test_dataset)
    preds =
    predictions.predictions.argmax(axis=1)

    from sklearn.metrics import accuracy_score,
    classification_report

    print("Accuracy:", accuracy_score(y_test,
    preds))
print(classification_report(y_test, preds))
```

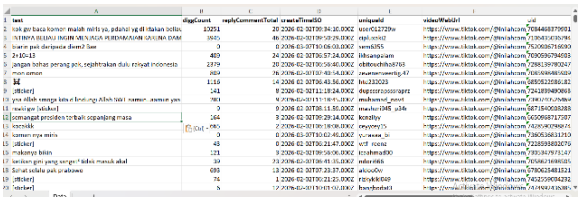
4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan dan Deskripsi Data

Penelitian ini menggunakan data komentar dari media sosial X dan TikTok yang berkaitan dengan konflik Iran–Israel. Data dikumpulkan melalui teknik crawling menggunakan API berbasis Python untuk X dan Apify Web Scraper untuk TikTok. Pada dataset X, atribut yang digunakan dalam penelitian ini adalah full_text yang berisi teks komentar pengguna. Contoh hasil crawling data dari platform X ditunjukkan pada gambar berikut:



Gambar 2. Hasil Crawling Dataset Komentar di X



Gambar 3. Hasil Crawling Dataset Komentar di TikTok

Total data yang diperoleh sebanyak 4.671 komentar, terdiri dari 1.671 komentar dari X dan 3.000 komentar dari TikTok. Dataset ini memiliki atribut utama berupa teks komentar yang digunakan sebagai input analisis.

4.2. Prapemrosesan Data

Tahap prapemrosesan dilakukan untuk meningkatkan kualitas data sebelum analisis menggunakan Natural Language Processing. Proses ini meliputi cleaning, case folding, dan tokenizing. Cleaning bertujuan menghilangkan noise seperti emoji, simbol, dan URL. Case folding menyeragamkan huruf menjadi lowercase, sedangkan tokenizing memecah kalimat menjadi unit kata. Hasil prapemrosesan menunjukkan bahwa data menjadi lebih bersih, terstruktur, dan siap untuk analisis lanjutan.

Tabel 1. Contoh Hasil Prapemrosesan Data

No	Komentar Asli	Hasil Cleaning	Token
1	@user iran kuat banget 💪🔥	iran kuat banget	[iran, kuat, banget]
2	perang ini bikin takut banget 😨	perang ini bikin takut banget	[perang, ini, bikin, takut, banget]
3	semoga dunia damai tanpa perang	semoga dunia damai tanpa perang	[semoga, dunia, damai, tanpa, perang]
4	#iran vs #israel makin panas!!!	iran vs israel makin panas	[iran, vs, israel, makin, panas]
5	kasian korban perang, semoga cepat selesai	kasian korban perang semoga cepat selesai	[kasian, korban, perang, semoga, cepat, selesai]

4.3. Pelabelan Sentimen

Pelabelan dilakukan menggunakan pendekatan lexicon-based dengan tiga kategori sentimen: positif, negatif, dan netral.

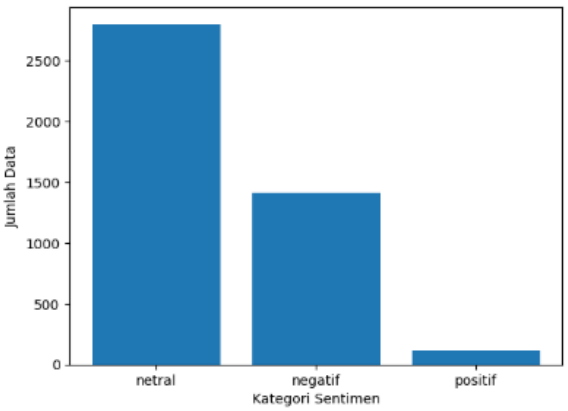
Tabel 2. Distribusi Sentimen

Sentimen	Jumlah Data
Positif	247
Netral	3066
Negatif	1358
Total	**4671**

Hasil tersebut menunjukkan dominasi sentimen netral sebanyak 3.066 data, diikuti negatif sebanyak 1.358 data, dan positif sebanyak 247 data. Dominasi sentimen netral menunjukkan bahwa sebagian besar komentar bersifat informatif, sedangkan tingginya sentimen negatif dipengaruhi oleh konteks konflik yang memicu kekhawatiran publik.

4.4. Klasifikasi Menggunakan IndoBERT

Proses klasifikasi dilakukan menggunakan model IndoBERT dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Model melakukan klasifikasi berdasarkan probabilitas tertinggi pada tiap kategori. Hasil klasifikasi menunjukkan bahwa komentar bernuansa konflik cenderung dikategorikan sebagai negatif, sedangkan komentar ringan atau informatif diklasifikasikan sebagai netral.



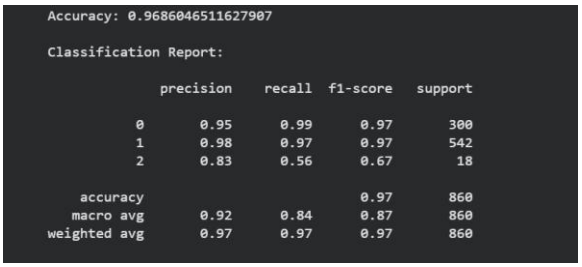
Gambar 4. Grafik Distribusi Sentimen Hasil Klasifikasi IndoBERT

Berdasarkan Gambar 4, distribusi hasil klasifikasi didominasi oleh sentimen netral, diikuti sentimen negatif dan positif. Dominasi sentimen netral menunjukkan bahwa sebagian besar komentar bersifat informatif, sedangkan tingginya sentimen negatif dipengaruhi oleh respons masyarakat terhadap konflik Iran–Israel.

4.5. Evaluasi Model IndoBERT

Evaluasi model IndoBERT dilakukan menggunakan data uji sebanyak 934 data yang merupakan 20% dari total dataset. Berdasarkan hasil pengujian, diperoleh nilai accuracy sebesar 96,86%, yang menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar.

Selain itu, performa model juga diukur menggunakan precision, recall, dan F1-score untuk masing-masing kategori sentimen. Hasil evaluasi tersebut ditampilkan pada Gambar berikut:

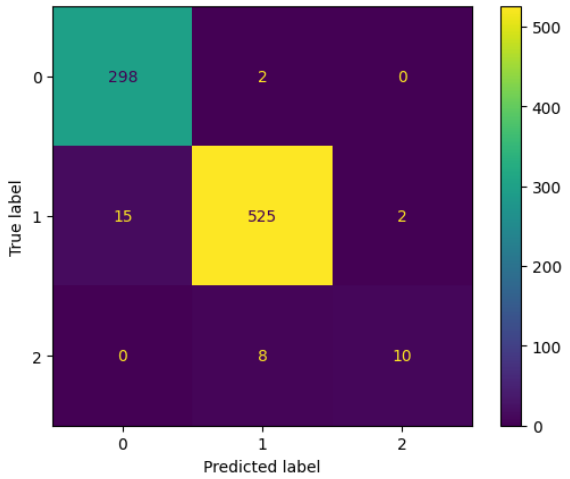


```
Accuracy: 0.9686046511627907

Classification Report:

              precision    recall  f1-score   support

           0       0.95      0.99      0.97       300
           1       0.98      0.97      0.97       542
           2       0.83      0.56      0.67        18

    accuracy                           0.97       860
   macro avg       0.92      0.84      0.87       860
weighted avg       0.97      0.97      0.97       860
```

Gambar 5. Classification Report

Berdasarkan Gambar 5, model IndoBERT menunjukkan performa yang sangat baik pada kategori sentimen negatif dan netral dengan nilai precision, recall, dan F1-score yang tinggi. Namun, pada kategori sentimen positif, performa model relatif lebih rendah dengan nilai recall sebesar 0,56 dan F1-score sebesar 0,67.

Untuk melihat distribusi hasil prediksi model, digunakan confusion matrix yang ditampilkan pada gambar berikut:



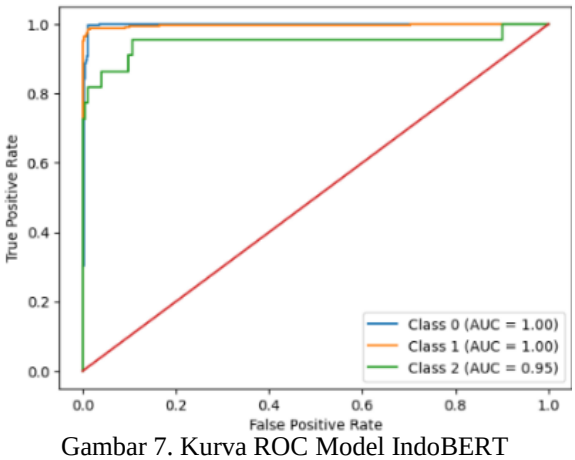
Gambar 6. Confusion Matrix Model IndoBERT

Berdasarkan Gambar 6, sebagian besar data berhasil diklasifikasikan dengan benar, yang ditunjukkan oleh nilai diagonal utama yang tinggi. Namun, masih terdapat kesalahan klasifikasi, terutama pada data sentimen positif yang sering diprediksi sebagai netral.

Hal ini menunjukkan bahwa model lebih cenderung mengklasifikasikan data ke kelas yang lebih dominan, yaitu netral, akibat ketidakseimbangan data.

Selain itu, hasil evaluasi juga dianalisis menggunakan kurva ROC yang ditampilkan pada gambar berikut:

Berdasarkan Gambar 7, model menunjukkan performa yang sangat baik dengan nilai AUC yang mendekati 1 pada hampir semua kelas. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan antar kelas sentimen.



Gambar 7. Kurva ROC Model IndoBERT

Namun, pada kelas sentimen positif, nilai AUC sedikit lebih rendah dibandingkan kelas lainnya. Hal ini kembali menunjukkan bahwa jumlah data yang tidak seimbang mempengaruhi kemampuan model dalam mengenali kelas tersebut.

Secara keseluruhan, model IndoBERT mampu memberikan hasil klasifikasi yang akurat, namun performanya masih dipengaruhi oleh distribusi data yang tidak seimbang

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa seluruh tahapan analisis sentimen opini publik terhadap berita trending di media sosial telah berhasil dilaksanakan secara sistematis, mulai dari pengumpulan data, prapemrosesan, pelabelan, hingga klasifikasi menggunakan model IndoBERT. Model IndoBERT terbukti mampu mengolah dan mengklasifikasikan data teks secara efektif dengan tingkat akurasi yang sangat baik sebesar 96,86%, sehingga mampu memahami konteks bahasa Indonesia, termasuk teks informal yang umum ditemukan di media sosial. Hasil analisis menunjukkan bahwa opini publik didominasi oleh sentimen netral, diikuti oleh sentimen negatif, dan sentimen positif dalam jumlah paling sedikit, yang mencerminkan bahwa sebagian besar pengguna cenderung menyampaikan informasi tanpa ekspresi emosi yang kuat, sementara sentimen negatif muncul sebagai respons terhadap isu konflik yang dibahas. Namun demikian, performa model pada kelas sentimen positif masih relatif lebih rendah akibat ketidakseimbangan jumlah data, sehingga mempengaruhi kemampuan model dalam mengenali pola sentimen positif secara optimal.

Berdasarkan hasil penelitian, disarankan agar penelitian selanjutnya menggunakan dataset dengan distribusi sentimen yang lebih seimbang untuk mengatasi keterbatasan model dalam mengenali kelas dengan jumlah data yang sedikit, khususnya sentimen positif. Selain itu, proses pelabelan data perlu dikembangkan dengan pendekatan yang lebih kontekstual, misalnya melalui kombinasi metode lexicon-based dan pelabelan manual, guna

meningkatkan kualitas data latih dan mengurangi kesalahan klasifikasi. Penelitian berikutnya juga dapat melakukan perbandingan dengan model lain seperti LSTM atau model transformer selain IndoBERT untuk memperoleh performa terbaik dalam analisis sentimen. Lebih lanjut, pengembangan analisis dapat diarahkan pada pendekatan yang lebih spesifik, seperti aspect-based sentiment analysis atau analisis emosi agar menghasilkan informasi yang lebih mendalam. Terakhir, disarankan untuk memperluas objek penelitian dengan berbagai topik dan sumber data yang berbeda sehingga hasil penelitian menjadi lebih general dan representatif terhadap opini publik secara luas.

DAFTAR PUSTAKA

[1] Danur Lambang Pristiandaru, "Iran Sepakat Gencatan Senjata dengan AS, tapi Tegaskan Perang Belum Berakhir," Kompas.com. [Online]. Available: https://internasional.kompas.com/read/2026/04/08/075100070/iran-sepakat-gencatan-senjata-dengan-as-tapi-tegaskan-perang-belum?utm_source=Various&utm_medium=Referral&utm_campaign=AIML_Widget_Desktop

[2] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.

[3] M. R. Manoppo, "Analisis Sentimen Publik di Media Sosial terhadap Kenaikan PPN 12% di Indonesia Menggunakan IndoBERT," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 4, no. 2, pp. 152–163, 2025, doi: 10.69916/jkbti.v4i2.322.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[5] D. Purnamasari, *Pengantar Metode Analisis Sentimen*. 2023.

[6] W. Wongso, D. S. Setiawan, S. Limcorn, and A. Joyoadikusumo, "NusaBERT: Teaching IndoBERT to be Multilingual and Multicultural," in *Proceedings of COLING*, 2025, pp. 10–26.

[7] D. K. Sumartha, "Opini Publik Terhadap Kebijakan Kenaikan Ukt Di Era Pemerintahan," vol. 13, no. 3, 2025.

[8] V. Foswanto, E. Sulistianingsih, and H. Perdana, "Implementasi Web Scraping untuk Analisis Ulasan Film KKN di Desa Penari Menggunakan Naive Bayes Classifier," *Equator J. Math. Stat.*, vol. 3, no. 1, pp. 1–11, 2024.

[9] D. I. Putri, A. N. Alfian, M. Y. Putra, and P. D. Mulyo, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," *J. Appl. Informatics Comput.*, vol. 8, no. 1, pp. 7–12, 2024, doi: 10.30871/jaic.v8i1.7440.

[10] F. Sugandi, "Analisis Sentiment Masyarakat Indonesia Pada Media Sosial Terhadap Isu Ijazah Palsu Mantan Presiden Menggunakan Algoritma Berbasis Transformer (BERT)," *J. Penelit. Sains dan Sos.*, vol. 8, no. 3, pp. 4762–4768, 2025, doi: 10.54314/jssr.v813.4209.

[11] Siti Rihastuti and Afnan Rosyidi, "Analisis Sentimen Pengguna Tiktok Tentang Progres Pembangunan Ikn Dengan Metode Random Forest," *J. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 19–23, 2025, doi: 10.54840/jcstech.v5i1.345.

[12] A. Averina, H. Hadi, and J. Siswantoro, "Analisis Sentimen Multi-Kelas untuk Film Berbasis Teks Ulasan Menggunakan Model Regresi Logistik," *Teknika*, vol. 11, no. 2, pp. 123–128, 2022, doi: 10.34148/teknika.v11i2.461.