

Analisis Perbedaan Pendapat Warganet Indonesia dan Internasional Terkait Aksi Boikot Produk Pendukung Israel

Arry Avorizano, Adityo Fajar Nugroho

Fakultas Teknologi Industri dan Informatika, Universitas Muhammadiyah Prof. DR. HAMKA, Indonesia

Email: ¹avorizano@uhamka.ac.id, ²adityofajarn.afn@gmail.com

Article Info

Received: February 12, 2025

Accepted: March 18, 2025

Published: September 31, 2025

ABSTRAK

Konflik antara Israel dan Palestina kini menjadi sorotan secara global. Karena sejumlah produk perusahaan, secara terang-terangan memberikan donasi berupa uang dan barang ke Israel. Hal ini membuat masyarakat marah terhadap Israel serta produk yang mendukungnya, sehingga menyerukan aksi boikot produk pendukung Israel di media sosial. Oleh karena itu, peneliti melakukan analisis sentimen terkait aksi boikot produk pendukung Israel dari media sosial X dan YouTube menggunakan algoritma Naïve Bayes dengan ketiga modelnya, yaitu *Multinomial*, *Bernoulli*, dan *Gaussian* pada warganet Indonesia dan internasional. Hasilnya, pada dataset internasional mendapatkan nilai sentimen positif sebesar 79.79% dengan jumlah 600 data yang mendukung aksi boikot, sedangkan pada kelas negatif diperoleh 20.21% dengan jumlah 152 data yang menolak aksi boikot. Kemudian pada dataset Indonesia memiliki nilai sentimen positif sebesar 51.78% dengan jumlah 1312 data yang mendukung aksi boikot, sedangkan pada kelas negatif diperoleh 48.22% dengan jumlah 1222 data yang menolak aksi boikot. Jadi, di Indonesia ataupun negara lainnya, mayoritas warganet menyerukan aksi boikot produk pendukung Israel melalui platform media sosial X dan YouTube. Saat dilakukan evaluasi model, model *BernoulliNB* mendapatkan nilai akurasi paling tinggi pada kedua dataset. Pada dataset Indonesia, *BernoulliNB* mencapai nilai akurasi hingga 89%. Kemudian pada dataset internasional, mendapatkan nilai akurasi sebesar 76%.

Kata kunci : analisis sentimen, boikot, israel, palestina, naïve bayes, x, youtube

ABSTRACT

The conflict between Israel and Palestine is now in the global spotlight. Because a number of company products openly donate money and goods to Israel. This has made people angry at Israel and the products that support it, thus calling for a boycott of products supporting Israel on social media. Therefore, researchers conducted sentiment analysis related to the boycott of products supporting Israel from social media X and YouTube using the Naïve Bayes algorithm with its three models, namely Multinomial, Bernoulli, and Gaussian on Indonesian and international citizens. As a result, the international dataset obtained a positive sentiment value of 79.79% with 600 data supporting the boycott, while the negative class obtained 20.21% with 152 data rejecting the boycott. Then the Indonesian dataset has a positive sentiment value of 51.78% with 1312 data supporting the boycott, while in the negative class 48.22% is obtained with 1222 data rejecting the boycott. So, in Indonesia and other countries, the majority of netizens call for a boycott of products supporting Israel through social media platforms X and YouTube. When evaluating the model, the BernoulliNB model obtained the highest accuracy value on both datasets. On the Indonesian dataset, BernoulliNB achieved an accuracy value of up to 89%. Then on the international dataset, it gets an accuracy value of 76%.

Keywords : sentiment analysis, boycott, israel, palestine, naïve bayes, x, youtube

1. PENDAHULUAN

Konflik antara Israel dan Palestina hingga kini belum menemukan penyelesaian, terus berlanjut tanpa titik temu yang jelas. Klaim Yahudi terhadap tanah Palestina menciptakan suatu ketidaksetaraan yang besar dan menimbulkan ancaman kedua belah pihak. Israel terus melakukan upaya perluasan permukiman di tanah Palestina tanpa henti. Palestina terus menderita dan hanya bisa melawan semampunya, tetapi pada akhirnya, sedikit demi sedikit tanah mereka diambil alih untuk dijadikan pemukiman Yahudi (Mudore, 2019). Serangan Israel yang berlangsung sejak 7 Oktober 2023 terus memakan korban. Menurut Palestinian Central Bureau of Statistics (PCBS) dalam rentang waktu 7 Oktober 2023 - 29 Desember 2024, mengakibatkan jumlah total korban jiwa Palestina mencapai 46.349 orang meninggal, 114.639 orang terluka dan lebih dari 2 juta orang terlanter (PCBS, 2024).

Puncak kemarahan masyarakat terjadi ketika Israel meluncurkan serangan menuju tempat-tempat yang tidak seharusnya menjadi sasaran. Beberapa diantaranya adalah tempat pengungsian, rumah sakit, dan tempat ibadah (Akbar, 2023). Sejak konflik antara Hamas dan Israel di wilayah Gaza berkembang menjadi tragedi kemanusiaan, muncul seruan boikot dari konsumen dalam negeri terhadap produk atau merek yang terhubung dengan negara Israel. Hal ini terjadi karena sejumlah produk, baik asing maupun lokal, secara terang-terangan mendukung tindakan Israel, termasuk dengan memberikan donasi berupa uang dan barang. Hal tersebut membuat masyarakat semakin marah terhadap Israel serta produk yang mendukungnya (Faizin, 2024). Akibatnya, melalui media sosial para warganet mulai gencar menyerukan aksi boikot pada produk pendukung Israel. Aksi itu disebut dengan Boycott, Divestment and Sanctions (BDS) yaitu upaya untuk menekan Israel secara ekonomi dengan tujuan mendorong pemberian hak setara kepada Palestina (Rosseno, 2023).

Media sosial menjadi tempat yang paling mudah untuk menyerukan pendapat, terutama pada platform X dan YouTube. Oleh karena itu, perlu dilakukan analisis sentimen warganet pada media sosial X dan YouTube terkait aksi boikot produk pendukung Israel. Hasil ini bertujuan untuk memperlihatkan seberapa banyak sentimen positif dan negatif terkait aksi boikot produk pendukung Israel di Indonesia maupun secara global. Dengan ketentuan, apabila setuju untuk boikot, maka teks tersebut masuk ke sentimen positif dan untuk yang menolak boikot, maka masuk ke sentimen negatif.

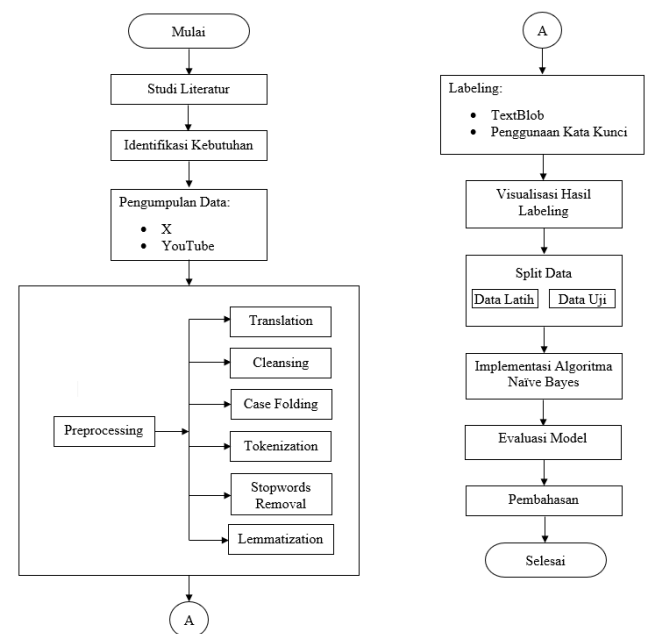
Algoritma Naïve Bayes memiliki 3 model, yaitu *Multinomial*, *Bernoulli*, dan *Gaussian*. Dari ketiga model tersebut, model *Multinomial* paling sering digunakan dalam analisis sentimen. Seperti penelitian yang dilakukan oleh Mandasari et al, hanya dengan menggunakan model *Multinomial* Naïve Bayes saja, dan menghasilkan nilai akurasi sebesar 86,57% (Mandasari et al., 2022). Oleh karena itu, peneliti menggunakan metode algoritma machine learning Naïve Bayes dengan menerapkan ketiga modelnya yaitu, *Multinomial*, *Bernoulli*, dan *Gaussian* dalam penelitian kali ini. Dengan tujuan, untuk memperlihatkan model mana yang paling baik dalam hal evaluasi model. Serta, pada penelitian yang telah dilakukan oleh Az-haari et al, hanya menggunakan platform media

sosial Twitter saja untuk pengumpulan datanya dan hanya berdasarkan kata kunci berbentuk *hashtag* (*#BoikotStarbucks* dan *#BoikotProIsrael*) dan menghasilkan akurasi sebesar 69% (Az-haari et al., 2024). Sehingga untuk menciptakan kebaruan dan variasi data, peneliti menggunakan dua platform media sosial, yaitu X dan juga YouTube untuk pengumpulan data.

2. METODE PENELITIAN

2.1. ALUR PENELITIAN

Pada tahap ini akan digambarkan diagram alur, guna untuk melihat tahapan penelitian yang dilakukan oleh peneliti. Diagram tersebut dibuat sebagai pedoman penelitian yang dapat dilakukan secara struktur sehingga dapat menghasilkan sistem yang mumpuni, seperti pada Gambar 1.



Gambar 1. Alur Penelitian

2.2. PENJELASAN DIAGRAM ALUR

- Studi Literatur**
Studi literatur merupakan serangkaian langkah yang mencakup pengumpulan data dari sumber pustaka, proses membaca dan mencatat informasi, serta pengolahan bahan-bahan yang digunakan dalam penelitian (Wiradi, 2022).
- Identifikasi Kebutuhan**
Identifikasi kebutuhan merupakan tahapan yang mendukung dalam proses tahap penelitian. Agar dapat melanjutkan ke tahap selanjutnya, persiapan yang harus dipenuhi berupa perangkat lunak dan perangkat keras komputer sebagai media untuk membantu proses penelitian.
- Pengumpulan Data (Crawling)**
Crawling adalah proses pengambilan data dari situs web atau platform media sosial seperti Facebook, X, YouTube, dan lainnya. Proses *crawling* memungkinkan pengumpulan data

secara otomatis dan efisien dalam jumlah besar, membantu memahami struktur informasi di internet, dan menghasilkan dataset yang dapat digunakan untuk berbagai tujuan analisis (Akbar, 2023). Pengumpulan data dilakukan dengan teknik *crawling* pada platform media sosial X menggunakan kata kunci 'boycott israel product' dengan rentang waktu dari 1 September 2023 hingga 30 November 2024 dan hanya mengambil *tweet* dari pengguna yang menggunakan bahasa Inggris dengan bantuan *Tweet-harvest*. *Tweet-harvest* adalah *library open-source* pada bahasa pemrograman Python, yang memungkinkan pengguna untuk mengumpulkan *tweet* dengan kata kunci tertentu dan jumlah *tweet* yang ditentukan (Rafli Ghufon et al., 2023). Kemudian peneliti juga mengumpulkan data dari video YouTube yang membahas mengenai aksi boikot produk pendukung Israel pada video berbahasa Indonesia, dengan menggunakan layanan API dari google yaitu YouTube Data API v3 dan modul *googleapiclient* dengan bahasa pemrograman Python (Rofiqi & Akbar, 2024)

d. *Preprocessing*

Tahap ini merupakan langkah awal yang sangat penting untuk menyiapkan data sebelum diproses. Data hasil dari *crawling* tergolong masih sangat mentah, ada saja kendala yang terjadi pada data, seperti terdapat data yang sama, nilai yang hilang, serta simbol yang tidak penting. Oleh karena itu, perlu dilakukan *preprocessing* yang bertujuan untuk membersihkan dan mengubah data agar menjadi format yang bisa dibaca oleh algoritma. Ada beberapa langkah yang akan dilakukan oleh peneliti pada tahap *preprocessing*, yaitu:

1. Translation: Mengubah bahasa satu ke bahasa lainnya. Peneliti menggunakan bantuan *library googletrans* untuk mengubah seluruh bahasa ke dalam Bahasa Inggris.
2. Cleansing: Pada tahap ini, dilakukan pembersihan data dengan cara menghapus simbol, emoji, *URL*, nilai yang hilang, duplikasi data dan menghapus semua karakter selain huruf. Hal ini harus dilakukan, karena komponen-komponen tersebut tidak diperlukan dalam pengolahan data serta bertujuan untuk mencegah agar proses analisis tidak terganggu.
3. Case Folding: Pada tahap ini peneliti mengubah semua huruf 'a' sampai 'z' dan huruf depan ataupun belakang menjadi huruf kecil (*lowercase*) agar masuk ke dalam satu format yang konsisten.
4. Tokenization: Tahap ini dilakukan untuk memisahkan kalimat serta paragraf menjadi unit yang lebih kecil seperti kata. Setiap unit yang lebih kecil ini disebut dengan token, token bisa berupa huruf, angka, dan tanda baca.

5. Stopwords Removal: Langkah untuk melakukan penghapusan kata-kata yang sering muncul dalam bahasa (seperti 'other', 'only', 'will', dll) yang tidak relevan (*stopwords*) atau memberikan nilai tambah dalam analisis dan juga berfungsi untuk menyederhanakan teks.

6. Lemmatization: Tahap ini dilakukan untuk mengubah kata ke dalam bentuk dasar yang valid berdasarkan maknanya. Dengan menggunakan pustaka Python yang bernama *SpaCy* (Supriyanto et al., 2023).

e. *Labeling*

Labeling adalah proses untuk memberikan tanda atau label pada data dengan nilai sentimen positif atau negatif. Pelabelan ini dapat dilakukan dengan berbagai metode, di antaranya adalah dengan cara manual yaitu menginput sendiri nilai sentimen yang terkandung pada teks, dengan Python dapat juga dilakukan menggunakan library *TextBlob* (Arham, 2023). Pada tahap ini dua kali dilakukan pelabelan, yang pertama dengan *labeling* otomatis menggunakan *TextBlob*. Kemudian pemberian label berdasarkan kata kunci, dengan ketentuan, opini yang mendukung aksi boikot terhadap produk-produk yang mendukung Israel dikelompokkan sebagai sentimen positif (*boycott israel, keep boycott*), sementara opini yang menolak aksi boikot dimasukkan dalam kelompok sentimen negatif (*do not boycott, never boycott*).

f. Visualisasi Data

Visualisasi data merupakan representasi grafis dari informasi dan data. Dalam data visualization, elemen-elemen visual seperti diagram dan grafik digunakan untuk menyajikan data secara lebih mudah dipahami (Madyatmadja et al., 2021). Pada tahap ini peneliti menampilkan bentuk visualisasi menggunakan diagram batang untuk melihat perbandingan jumlah sentimen positif dan sentimen negatif pada dataset internasional dan Indonesia.

g. *Split Data*

Split Data merupakan pembagian data menjadi dua segmen, yaitu data latih (*train*) dan data uji (*test*). Dengan nilai pembagian 80% dari data akan dialokasikan untuk set data latih, dan sisanya 20% digunakan untuk set data uji. Data *train* digunakan untuk melatih model, sedangkan data *test* digunakan untuk mengukur seberapa baik kinerja model pada data yang belum pernah dilihat sebelumnya.

h. Implementasi Algoritma Naive Bayes

Pada tahap ini peneliti melakukan implementasi algoritma Naive Bayes menggunakan 3 model yaitu *Multinomial*, *Bernoulli*, dan *Gaussian*. s Algoritma Naive Bayes Classifier tergolong dalam kelompok algoritma yang digunakan dalam teknik klasifikasi. Metode ini melibatkan perkiraan probabilitas kejadian di masa depan berdasarkan pengalaman masa lalu, yang dikenal sebagai Teorema Bayes. Teorema ini digunakan

bersama dengan pendekatan Naïve, di mana diasumsikan bahwa atribut-atribut tersebut bersifat independen satu sama lain. Dalam konteks klasifikasi Naïve Bayes, diasumsikan bahwa keberadaan atau ketiadaan ciri tertentu dari suatu kelas tidak memiliki hubungan dengan ciri dari kelas lainnya (Bustami, 2019).

$$P(B) = \frac{P(B|A).P(A)}{P(B)} \quad (1)$$

- i. Evaluasi Model
Evaluasi model menggunakan *confusion matrix*. *Confusion Matrix* adalah metode evaluasi berbasis matriks yang digunakan untuk mengukur performa model klasifikasi. Metode ini membandingkan hasil prediksi yang diberikan oleh model dengan nilai aktual, sehingga memberikan gambaran tentang seberapa baik model tersebut dalam melakukan klasifikasi (Purnomo et al., 2024).

Tabel 1. Confusion Matrix

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

Kemudian, berdasarkan Tabel 1. Istilah tersebut digunakan untuk menghitung nilai akurasi, *precision*, *recall*, dan *f1-score* dengan rumus di bawah ini.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

- j. Pembahasan
Setelah semua proses dilakukan, peneliti merangkum hasil-hasil yang diperoleh ke dalam tahap pembahasan. Pada tahap ini, peneliti akan menjelaskan hasil perbandingan klasifikasi positif atau negatif pada dataset internasional dan Indonesia, serta membahas akurasi yang diperoleh dari ketiga model Naïve Bayes yang telah diuji. Sehingga, dapat dilihat model mana yang memberikan hasil terbaik.

3. HASIL DAN PEMBAHASAN

3.1. PENGUMPULAN DATA

3.1.1. Pengumpulan Data Platform X

Pada tahap ini, peneliti mengumpulkan data dengan teknik *Crawling* pada platform X menggunakan *tweet-harvest*. Peneliti menggunakan tiga kata kunci, yaitu “boycott israel product”, dengan rentang waktu dari 1 September 2023 hingga 30 November 2024 dan hanya mengambil tweet dari pengguna yang menggunakan bahasa Inggris, seperti yang terlihat pada Gambar 2.

	A	B	C	D	E	F
1	conversat	on_id_str,"created_at","favorite_count","full_text","ic				
2	1807049808060059953,"Sat Jun 29 13:53:42 +0000 2024","82					
3	1807020024269574361,"Sat Jun 29 11:55:21 +0000 2024","1"					
4	1806963365258183078,"Sat Jun 29 08:10:12 +0000 2024","0"					
5	1806663192900223127,"Fri Jun 28 12:39:38 +0000 2024","0",					

Gambar 2. Crawling Data X

3.1.2. Pengumpulan Data Platform YouTube

Dataset Indonesia dikumpulkan melalui media sosial YouTube, karena pengguna Indonesia aktif berkomentar di platform tersebut. Pengumpulan data komentar warganet Indonesia menggunakan layanan API dari google yaitu YouTube Data API v3 dan modul *googleapiclient* dengan bahasa pemrograman Python. Data yang dikumpulkan berasal dari salah satu video YouTube yang membahas aksi boikot produk pendukung Israel. Dari video tersebut, tersebut peneliti mendapatkan 3245 data komentar YouTube, terlihat pada Gambar 3.

	A	B	C	D	E
1	publishedAt,authorDisplayName,full_text,likeCount				
2	2024-08-22T05:41:09Z,@RajibMuslim-eh1tj,Demi me				
3	2024-08-21T11:01:41Z,@FrankyTiwang,Oikot no way				
4	2024-08-20T11:00:18Z,@FrankyTiwang,Mui cuma ac				
5	2024-07-19T12:20:27Z,@AbaMinake,"Boikot HP aja,				

Gambar 3. Crawling Data YouTube

3.2. PREPROCESSING

Pada tahap ini, data hasil *Crawling* akan dilakukan proses pembersihan data menjadi data yang terstruktur dan efisien. Data yang tersimpan memiliki format CSV dan tergolong masih sangat mentah. Karena prosesnya hampir sama, peneliti hanya akan menampilkan proses preprocessing pada dataset internasional yang dikumpulkan dari platform X. Untuk menyempurnakan data ini di perlukan beberapa langkah, yaitu sebagai berikut:

a. Translation

Peneliti mengubah seluruh bahasa pada kedua dataset ke Bahasa Inggris sepenuhnya, untuk mempermudah proses pengolahan data. Peneliti menggunakan pustaka *googlettrans* untuk mengubah seluruh bahasa ke dalam Bahasa Inggris. Berikut ini merupakan contoh kalimat sebelum dan sesudah dilakukan *translation* ke dalam Bahasa Inggris, dapat dilihat pada Tabel 2.

Tabel 2. Tahap Translation

Sebelum Translation	Sesudah Translation
Hum bhi lumber one agency sy boycott kr chukay hen jo ky real israel ki product hy	We have also done the Lumber one agency and the product of real Israel
Kau nak suruh dia boycott Israel and at the same time suruh jugak boycott product dia just sebab dia nonMalaynonMuslim Pastu wonder where we go wrong hahahhah	You want to tell him to boycott Israel and at the same time tell also boycott products he is just because he is nonMalaynonMuslim surely wonder where we go wrong hahahhah

b. Cleansing

Pada tahap ini peneliti akan menghapus beberapa hal yang tidak dibutuhkan, seperti menghapus baris yang bernilai NaN, menghapus *retweet*, menghapus *hashtag*, menghapus *mention*, menghapus URL, menghapus karakter non-alfanumerik, menghapus spasi yang berlebih dan menghapus *tweet* duplikat. Pada Tabel 3 menampilkan perbandingan antara data *tweet* sebelum dan sesudah dilakukan proses *cleansing*. Pada kolom Sesudah Cleansing, terlihat bahwa data telah bersih dari simbol, emoji, URL, dan lain-lain.

Tabel 3. Tahap Cleansing

Sebelum Cleansing	Sesudah Cleansing
Every product bought from Israel supports their oppression of Palestinians. Boycott now and stand with Free Gaza! #ÜØ±ÜÊ_Ø²Ü https://t.co/KKipGEf6pF	Every product bought from Israel supports their oppression of Palestinians Boycott now and stand with Free Gaza
Sanction should be imposed on Israel. DONT BUY ISRAEL'S PRODUCT! #BoycottIsraelProduct #BanIsraelFromOlympics #BoycottIsraelProduct #BDS #Olympic #boycott #ban_Israel_from_Olympic #Murderers cannot beheroes https://t.co/C7MDDFWzyC	Sanction should be imposed on Israel DONT BUY ISRAELS PRODUCT

c. Case Folding

Pada tahap ini, peneliti melakukan *Case Folding* yang merupakan teknik untuk mengubah semua huruf besar menjadi huruf kecil (*lowercase*). Tujuannya untuk menyamaratakan penggunaan huruf kecil agar data lebih terstruktur dan menjaga konsistensi data. Berikut ini merupakan tabel perbandingan data sebelum dan sesudah dilakukan *Case Folding*, dapat dilihat pada Tabel 4.

Tabel 4. Tahap Case Folding

Sebelum Case Folding	Sesudah Case Folding
Every product bought from Israel supports their oppression of Palestinians Boycott now and stand with Free Gaza	every product bought from israel supports their oppression of palestinians boycott now and stand with free gaza
Sanction should be imposed on Israel DONT BUY ISRAELS PRODUCT	sanction should be imposed on israel dont buy israels product

d. Tokenization

Proses ini dilakukan untuk memisahkan kalimat serta paragraf menjadi unit yang lebih kecil dalam

bentuk token atau kata. Setiap unit yang lebih kecil ini disebut dengan token, token bisa berupa huruf, angka, dan tanda baca. Berikut ini merupakan tabel perbandingan data sebelum dan sesudah dilakukan *Tokenization*, dapat dilihat pada Tabel 5.

Tabel 5. Tahap Tokenization

Sebelum Tokenization	Sesudah Tokenization
every product bought from israel supports their oppression of palestinians boycott now and stand with free gaza	['every', 'product', 'bought', 'from', 'israel', 'supports', 'their', 'oppression', 'of', 'palestinians', 'boycott', 'now', 'and', 'stand', 'with', 'free', 'gaza']
sanction should be imposed on israel dont buy israels product	['sanction', 'should', 'be', 'imposed', 'on', 'israel', 'dont', 'buy', 'israels', 'product']

e. Stopwords Removal

Proses *Stopwords Removal* adalah langkah untuk melakukan penghapusan kata-kata yang sering muncul dalam bahasa (seperti “other”, “only”, “will”, dll) yang tidak relevan (*stopwords*) atau memberikan nilai tambah dalam analisis dan juga berfungsi untuk menyederhanakan teks. Berikut ini merupakan tabel perbandingan data sebelum dan sesudah dilakukan *Stopwords Removal*, dapat dilihat pada Tabel 6.

Tabel 6. Tahap Stopwords Removal

Sebelum Stopwords Removal	Sesudah Stopwords Removal
['every', 'product', 'bought', 'from', 'israel', 'supports', 'their', 'oppression', 'of', 'palestinians', 'boycott', 'now', 'and', 'stand', 'with', 'free', 'gaza']	['every', 'product', 'bought', 'israel', 'supports', 'oppression', 'palestinians', 'boycott', 'stand', 'free', 'gaza']
['sanction', 'should', 'be', 'imposed', 'on', 'israel', 'dont', 'buy', 'israels', 'product']	['sanction', 'imposed', 'israel', 'dont', 'buy', 'israels', 'product']

f. Lemmatization

Lemmatization mirip dengan proses *stemming* yang mana digunakan untuk mengubah kata ke dalam bentuk dasar yang valid berdasarkan maknanya. Peneliti menggunakan pustaka Python yang bernama *SpaCy* untuk melakukan proses *lemmatization*. *SpaCy* adalah *library* NLP (*Natural Language Processing*) yang populer dan efisien untuk melakukan berbagai tugas pengolahan teks, termasuk *lemmatization*. Berikut ini merupakan tabel perbandingan data sebelum dan sesudah dilakukan *Lemmatization*, dapat dilihat pada Tabel 7.

Tabel 7. Tahap Lemmatization

Sebelum Lemmatization	Sesudah Lemmatization
['every', 'product', 'bought', 'israel', 'supports', 'oppression', 'palestinians', 'boycott', 'stand', 'free', 'gaza']	['every', 'product', 'buy', 'israel', 'support', 'oppression', 'palestinians', 'boycott', 'stand', 'free', 'gaza']
['sanction', 'imposed', 'israel', 'dont', 'buy', 'israels', 'product']	['sanction', 'impose', 'israel', 'donot', 'buy', 'israel', 'product']

3.3. LABELING

Pada tahap ini, peneliti melakukan klasifikasi data dengan memberikan label terhadap data *tweet* menggunakan TextBlob. TextBlob adalah sebuah *library* Python yang dirancang untuk mempermudah proses analisis teks, seperti pemrosesan bahasa alami NLP (*Natural Language Processing*). TextBlob tidak secara langsung menentukan apakah kata itu positif atau negatif, melainkan memberikan label berdasarkan teks secara keseluruhan yang telah dilatih dengan Naïve Bayes. Pada penelitian ini, opini yang mendukung aksi boikot terhadap produk - produk yang mendukung Israel dikelompokkan sebagai sentimen positif, sementara opini yang menolak aksi boikot dimasukkan dalam kelompok sentimen negatif. Oleh karena itu, peneliti menambahkan pelabelan baru berdasarkan kata kunci untuk memperjelas proses klasifikasi. Pada Gambar 4. `positive_count` akan mencatat jumlah total kemunculan kata positif, yang berarti jika sebuah tweet mengandung beberapa kata kunci positif, masing-masing kata tersebut akan menambah poin. Begitu pula dengan `negative_count` untuk kata kunci negatif. Dengan cara ini, jumlah total kata kunci yang ditemukan akan berpengaruh pada penentuan sentimen.

```
def label_sentiment(comment):
    positive_keywords = ['always boycott', 'boycott israel', 'boycott isreali', 'free gaza']
    negative_keywords = ['stand israel', 'never boycott', 'confuse people']

    #mengitung jumlah kata kunci positif dan negatif
    positive_count = sum(comment.lower().count(keyword) for keyword in positive_keywords)
    negative_count = sum(comment.lower().count(keyword) for keyword in negative_keywords)

    #menentukan sentimen berdasarkan jumlah kata kunci yang paling banyak
    if positive_count > negative_count:
        return 'Positif'
    else:
        return 'Negatif'

data['klasifikasi'] = data['clean_text'].apply(label_sentiment)
data
```

Gambar 4. Proses Labeling Menggunakan Kata Kunci

Berikut ini merupakan beberapa hasil akhir dari *labeling* yang menghasilkan sentimen positif dan negatif yang dapat dilihat pada Tabel 8.

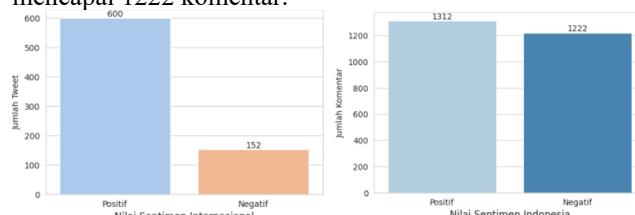
Tabel 8. Hasil Tahap Labeling

Sentimen Positif	Sentimen Negatif
alhamdulillah always boycott brand israel always awareness community cannot use israel product	never boycott israeli product find product israel wipe as put rubbish bin
israel scare boycott please people across world boycott every jewish product world also every product brand world product brand supporter genocide israel	never boycott israeli product jus buy thing like thing associate israel

3.4. VISUALISASI DATA

Pada nilai sentimen internasional yang terlihat di Gambar 10, data yang berhasil diproses berjumlah 752 tweet, dengan nilai sentimen positif mencapai 600 tweet, sedangkan untuk sentimen negatif hanya mencapai 152 tweet. Kemudian, pada nilai sentimen Indonesia yang terlihat di Gambar 5, data yang berhasil diproses berjumlah 2534 komentar, dengan nilai sentimen positif mencapai

1312 komentar, sedangkan untuk sentimen negatif mencapai 1222 komentar.



Gambar 5. Diagram Batang

3.5. SPLIT DATA

Pada tahap split data dilakukan pembagian data menjadi dua kelompok, yaitu data latih (X_{train} , y_{train}) dan data uji (X_{test} , y_{test}). Pembagian ini menggunakan metode `train_test_split` dari pustaka `sklearn.model.selection`. Menggunakan parameter `test_size=0.2` yang berarti 20% dari data akan dialokasikan untuk set data uji, dan sisanya 80% digunakan untuk set data latih. Dilakukan pembagian data pada kedua dataset dengan perbandingan 80:20. Pada dataset internasional, sebanyak 601 baris *tweet* dialokasikan untuk data latih dan 151 baris untuk data uji. Sedangkan pada dataset Indonesia, sebanyak 2027 baris komentar YouTube dialokasikan untuk data latih dan 507 baris untuk data uji.

3.6. IMPLEMENTASI ALGORITMA NAIVE BAYES

Sebelum menerapkan algoritma Naïve Bayes, dilakukan pengubahan teks menjadi representasi numerik berbasis TF-IDF yang berasal dari *library sklearn.feature_extraction.text*. Dapat dilihat pada Gambar 6.

```
#mengubah teks menjadi representasi numerik berbasis TF-IDF
vectorizer = TfidfVectorizer(stop_words='english', max_features=10000)
X_train_vectorized = vectorizer.fit_transform(X_train)
X_test_vectorized = vectorizer.transform(X_test)
```

Gambar 6. Vectorizer TF-IDF

Kemudian mulai mengimplementasikan algoritma Naïve Bayes menggunakan model *Multinomial*, *Bernoulli*, dan *Gaussian* untuk dibandingkan tingkat akurasi. *Alpha* dan *var_smoothing* berisi nilai 0.02 yang bertujuan untuk memberikan sedikit *smoothing* pada model. Tanpa *smoothing*, apabila terdapat kata yang tidak ada dalam data latih maka akan mengganggu prediksi model. Semakin besar nilai pada *alpha*, semakin kuat juga *smoothing*-nya. Dapat dilihat pada Gambar 7.


```
##model Multinomial Naive Bayes
mnb = MultinomialNB(alpha=0.02)
mnb.fit(X_train_vectorized, y_train)

#model Bernoulli Naive Bayes
bnb = BernoulliNB(alpha=0.02)
bnb.fit(X_train_vectorized, y_train)

#model Gaussian Naive Bayes
gnb = GaussianNB(var_smoothing=0.02)
gnb.fit(X_train_vectorized.toarray(), y_train)
```

Gambar 7. Implementasi Model Naive Bayes

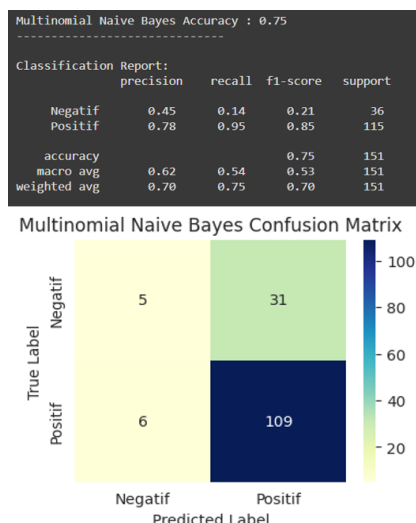
3.7. EVALUASI

Peneliti melakukan evaluasi model berdasarkan perhitungan yang akan ditampilkan dalam bentuk *classification report* dan tabel *confusion matrix*.

3.7.1. Evaluasi Model Dataset Internasional

a. Multinomial Naïve Bayes

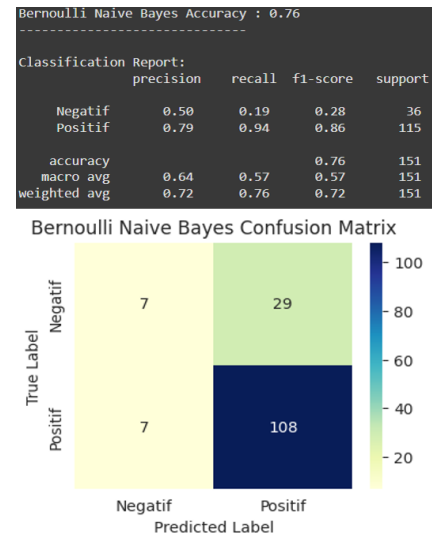
Multinomial Naïve Bayes mendapatkan nilai akurasi sebesar 75%. Pada kelas negatif, diperoleh *precision* sebesar 45%, *recall* 14%, dan *f1-score* 21%. Sedangkan pada kelas positif, diperoleh *precision* sebesar 78%, *recall* 95% dan *f1-score* 85%. Agar lebih jelas, dapat dilihat pada Gambar 8.



Gambar 8. Multinomial Naive Bayes Dataset Internasional

b. Bernoulli Naïve Bayes

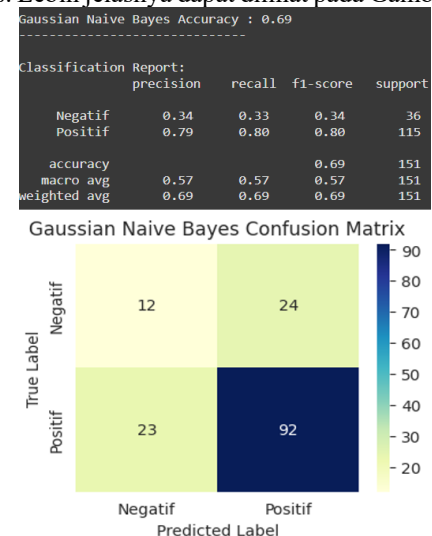
Bernoulli Naïve Bayes mendapatkan nilai akurasi sebesar 76%. Pada kelas negatif, diperoleh *precision* sebesar 50%, *recall* 19%, dan *f1-score* 28%. Sedangkan pada kelas positif, diperoleh *precision* sebesar 79%, *recall* 94% dan *f1-score* 86%. Agar lebih jelas, dapat dilihat pada Gambar 9.



Gambar 9. Bernoulli Naive Bayes Dataset Internasional

c. Gaussian Naïve Bayes

Gaussian Naïve Bayes mendapatkan nilai akurasi sebesar 69%. Pada kelas negatif, diperoleh *precision* sebesar 34%, *recall* 33%, dan *f1-score* 34%. Sedangkan pada kelas positif, diperoleh *precision* sebesar 79%, *recall* 80% dan *f1-score* 80%. Lebih jelasnya dapat dilihat pada Gambar 10.

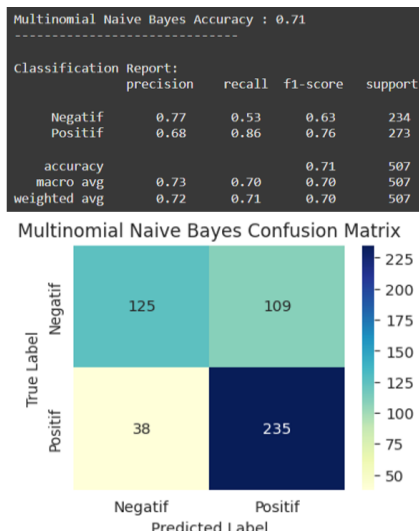


Gambar 10. Gaussian Naive Bayes Dataset Internasional

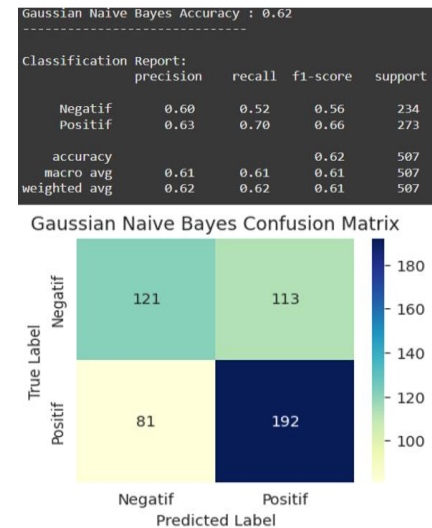
3.7.2. Evaluasi Model Dataset Indonesia

a. Multinomial Naïve Bayes

Multinomial Naïve Bayes mendapatkan nilai akurasi sebesar 71%. Pada kelas negatif, diperoleh *precision* sebesar 77%, *recall* 53%, dan *f1-score* 63%. Sedangkan pada kelas positif, diperoleh *precision* sebesar 68%, *recall* 86% dan *f1-score* 76%. Lebih jelasnya dapat dilihat pada Gambar 11.



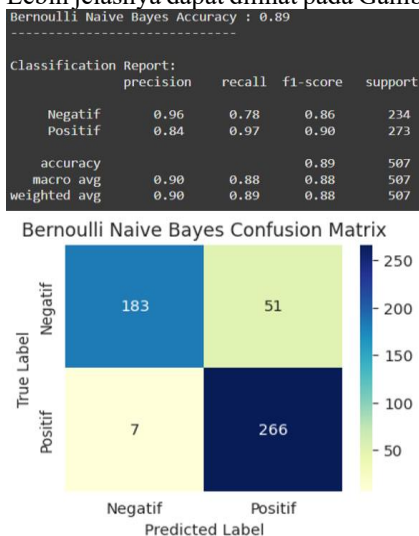
Gambar 11. Multinomial Naive Bayes Dataset Indonesia



Gambar 13. Gaussian Naive Bayes Dataset Indonesia

b. Bernoulli Naïve Bayes

Bernoulli Naïve Bayes mendapatkan nilai akurasi sebesar 89%. Pada kelas negatif, diperoleh *precision* sebesar 78%, *recall* 86%, dan *f1-score* 86%. Sedangkan pada kelas positif, diperoleh *precision* sebesar 84%, *recall* 97% dan *f1-score* 90%. Lebih jelasnya dapat dilihat pada Gambar 12.



Gambar 12. Bernoulli Naive Bayes Dataset Indonesia

c. Gaussian Naïve Bayes

Gaussian Naïve Bayes mendapatkan nilai akurasi sebesar 62%. Pada kelas negatif, diperoleh *precision* sebesar 60%, *recall* 52%, dan *f1-score* 56%. Sedangkan pada kelas positif, diperoleh *precision* sebesar 63%, *recall* 70% dan *f1-score* 66%. Lebih jelasnya dapat dilihat pada Gambar 13.

3.8. PEMBAHASAN

Berdasarkan analisis yang telah dilakukan oleh peneliti menggunakan algoritma Naïve Bayes, diperoleh hasil bahwa pada dataset internasional, kelas positif mencatatkan persentase sebesar 79,79% dengan jumlah 600 data, sementara kelas negatif memperoleh 20,21% dengan jumlah 152 data. Kemudian pada dataset Indonesia, kelas positif mencatatkan persentase sebesar 51,78% dengan jumlah 1312 data, sementara kelas negatif memperoleh 48,22% dengan jumlah 1222 data. Terlihat pada kedua dataset bahwa kelas positif lebih banyak dibandingkan dengan kelas negatif. Hal ini menunjukkan bahwa mayoritas warganet di X dan YouTube mendukung aksi boikot terhadap produk pendukung Israel.

Peneliti juga membandingkan dengan ketiga model Naïve Bayes, yaitu *MultinomialNB*, *BernoulliNB*, dan *GaussianNB*. Masing-masing model memiliki nilai akurasi yang berbeda. Pada dataset internasional, model *MultinomialNB* mendapatkan nilai akurasi sebesar 75%, *BernoulliNB* mendapatkan nilai akurasi sebesar 76%, dan *GaussianNB* mendapatkan nilai akurasi sebesar 69%. Kemudian pada dataset Indonesia, model *MultinomialNB* mendapatkan nilai akurasi sebesar 71%, *BernoulliNB* mendapatkan nilai akurasi sebesar 89%, dan *GaussianNB* mendapatkan nilai akurasi sebesar 62%. Dari kedua dataset, model *BernoulliNB* mendapatkan nilai akurasi paling tinggi diantara kedua model lainnya, karena memiliki model yang sederhana dan bekerja dengan baik ketika data terdiri dari kata-kata kunci yang sangat mendominasi pada kelas tertentu. Adanya kata kunci lebih mudah dipahami oleh model ini daripada model *MultinomialNB* yang berkerja sesuai dengan frekuensi kata. Sedangkan pada model *GaussianNB*, mendapatkan nilai akurasi paling rendah diantara kedua model lainnya, karena model ini mengasumsikan bahwa fitur-fitur mengikuti distribusi normal atau kontinu, sedangkan pada kedua dataset bersifat diskrit.

Perbedaan nilai rata-rata akurasi juga terlihat pada nilai *precision*, *recall*, dan *f1-score* di masing-masing dataset. Pada dataset internasional, model *BernoulliNB* memiliki nilai rata-rata yang paling baik, yaitu dengan *precision* mencapai 64%, *recall* 57%, dan *f1-score* 57%.

Model *MultinomialNB* mendapatkan nilai *precision* mencapai 62%, *recall* 54%, dan *f1-score* 53%. Model *GaussianNB* mendapatkan nilai *precision* mencapai 57%, *recall* 57%, dan *f1-score* 57%. Kemudian pada dataset Indonesia, model *BernoulliNB* memiliki nilai rata-rata yang paling baik juga, yaitu dengan *precision* mencapai 90%, *recall* 88%, dan *f1-score* 88%. Model *MultinomialNB* mendapatkan nilai *precision* mencapai 73%, *recall* 70%, dan *f1-score* 70%. Model *GaussianNB* mendapatkan nilai *precision* mencapai 61%, *recall* 61%, dan *f1-score* 61%.

Saat dilakukan evaluasi model pada dataset internasional, terlihat bahwa nilai *precision*, *recall*, dan *f1-score* untuk sentimen negatif memiliki akurasi yang sangat rendah jika dibandingkan dengan sentimen positif. Hal ini disebabkan oleh adanya ketidakseimbangan kelas, di mana jumlah data untuk kelas positif jauh lebih banyak dibandingkan dengan kelas negatif. Ketidakseimbangan kelas ini mempengaruhi performa model, karena model cenderung lebih baik dalam mendeteksi kelas yang memiliki lebih banyak sampel. Pengaruh terbesar kenapa *precision*, *recall*, dan *f1-score* untuk sentimen negatif sangat rendah, ialah dikarenakan proses pelabelan menggunakan kata kunci, sehingga data dikategorikan sebagai positif atau negatif berdasarkan keberadaan kata-kata tertentu. Jika kata kunci yang digunakan terlalu terbatas atau tidak bervariasi, hal ini membuat model jadi ketergantungan pada kata-kata tersebut.

Hasil akurasi yang diperoleh dari masing-masing dataset menunjukkan nilai yang berbeda. Dataset Indonesia memiliki tingkat akurasi yang lebih tinggi, yaitu mencapai 89% pada model *Bernoulli* Naïve Bayes, dibandingkan dengan dataset internasional yang hanya mencapai 76% akurasi pada model yang sama. Perbedaan nilai akurasi pada kedua dataset dapat disebabkan oleh jumlah data yang diproses. Dataset Indonesia terdiri dari 2534 komentar, sedangkan dataset internasional hanya berisi 752 baris *tweet*.

Jika hasil ini dibandingkan dengan penelitian sebelumnya yang dilakukan oleh Az-haari et al, maka penelitian yang dilakukan saat ini lebih baik dari penelitian sebelumnya dalam hal akurasi. Pada penelitian sebelumnya yang menggunakan model *Multinomial* Naïve Bayes mendapatkan nilai akurasi sebesar 69%, dengan sentimen berlabel positif sebanyak 130, negatif 151, dan netral 611 (Az-haari et al., 2024). Hasil akurasi pada penelitian sebelumnya lebih rendah dari pada penelitian yang dilakukan saat ini, hal ini dikarenakan *Bernoulli* Naïve Bayes cocok untuk analisis biner berdasarkan kata kunci, sederhana sederhana efisien untuk tugas klasifikasi sentimen seperti boikot ini. Perbedaan nilai akurasi ini juga disebabkan oleh adanya ketidakseimbangan kelas, pada penelitian sebelumnya, jumlah data untuk kelas netral jauh lebih banyak dibandingkan kelas positif dan negatif.

4. KESIMPULAN

Hasil yang diperoleh dari dataset internasional memiliki nilai sentimen positif sebesar 79.79% dengan jumlah 600 data *tweet* yang mendukung aksi boikot, sedangkan pada kelas negatif diperoleh 20.21% dengan jumlah 152 data *tweet* yang menolak aksi boikot.

Kemudian pada dataset Indonesia memiliki nilai sentimen positif sebesar 51.78% dengan jumlah 1312 data komentar yang mendukung aksi boikot, sedangkan pada kelas negatif diperoleh 48.22% dengan jumlah 1222 data komentar yang menolak aksi boikot. Jadi, di Indonesia ataupun negara lainnya, mayoritas warganet menyerukan aksi boikot produk pendukung Israel melalui platform media sosial X dan YouTube.

Saat mengimplementasikan ketiga model Naïve Bayes, model *Bernoulli* menunjukkan evaluasi yang paling baik pada kedua dataset dalam hal nilai akurasi, *precision*, *recall*, dan *f1-score*. Pada dataset Indonesia, model *Bernoulli* mendapatkan nilai akurasi sebesar 89%, rata-rata *precision* sebesar 90%, *recall* 88%, dan *f1-score* 88%. Sedangkan pada dataset internasional, model ini memperoleh nilai akurasi sebesar 76%, rata-rata *precision* sebesar 64%, *recall* 57%, dan *f1-score* 57%.

DAFTAR PUSTAKA

- Akbar, P. M. (2023). *Media Sosial Kian Riu dengan Seruan Boikot Produk Israel*. PT Republika Media Mandiri.
<https://www.republika.id/posts/47488/media-sosial-kian-riuh-dengan-seruan-boikot-produk-israel>
- Arham, A. (2023). *Labeling Sentimen Bahasa Indonesia Secara Otomatis Menggunakan Library TextBlob dan Googletrans* (Python).
<https://medium.com/@azriyanarham/labeling-sentimen-bahasa-indonesia-secara-otomatis-79764f23b016>
- Az-haari, N. F., Juardi, D., & Jamaludin, A. (2024). Analisis Sentimen Terhadap Boikot Brand Pro-Israel Pada Twitter Menggunakan Metode Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 4256–4261. <https://doi.org/10.36040/jati.v8i3.9888>
- Bustami. (2019). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi pada BPR Pantura. *TECHSI: Jurnal Penelitian Teknik Informatika*.
<https://repository.nusamandiri.ac.id/index.php/repo/viewitem/13890>
- Faizin, N. (2024). *Pemboikotan Produk Israel dalam Pandangan Hukum Islam Nur Faizin*. 73–84.
- Madyatmadja, E. D., Kusumawati, L., Jamil, S. P., Kusumawardhana, W., Informasi, S., & Nusantara, U. B. (2021). Infotech: journal of technology information. *Raden Ario Damar*, 7(1), 55–62.
- Mandasari, S., Hayadi, B. H., & Gunawan, R. (2022). Analisis Sentimen Pengguna Transportasi Online Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 5(2), 118.
<https://doi.org/10.53513/jsk.v5i2.5635>
- Mudore, S. B. (2019). Peran Diplomasi Indonesia Dalam Konflik Israel-Palestina. *Jurnal CMES*, 12(2), 170.
<https://doi.org/10.20961/cmcs.12.2.37891>
- PCBS. (2024). PCBS. Palestinian Central Bureau of Statistics. <https://www.pcbs.gov.ps/default.aspx>
- Purnomo, G., Susanto, T., Komputer, F. I., & Yogyakarta, U. A. (2024). *Implementasi Algoritma Support Vector*

- Machine (SVM) pada Analisis Sentimen Opini Publik Tentang Larangan Penggunaan Obat Sirup bagi Kesehatan Ginjal Implementation of The Support Vector Machine (SVM) Algorithm on Sentiment Analysis of Public Opinion o. 13, 2332–2347.*
- Rafli Ghufro, M., Farrih, M., Arsyada, M., Lukman, M. R., Azhar, Y., Putra, H., Rakhmawati, N. A., & Kunci, K. (2023). Analisis Sentimen Pengguna Twitter Terhadap Pemilu 2024 Berbasis Model XLM-T. *Information Adn Technology*, 204, 307–315.
- Rofiqi, L., & Akbar, M. (2024). Analisis Sentimen Terkait RUU Perampasan Aset dengan Support Vector Machine. *JEKIN - Jurnal Teknik Informatika*, 4(3), 529–538. <https://doi.org/10.58794/jekin.v4i3.824>
- Roseno, A. N. (2023). *RI & Banyak Negara Boikot Produk Israel, Apa Efeknya?* CNBC Indonesia. [https://www.cnbcindonesia.com/news/20231114170902-4-488970/ri-banyak-negara-boikot-produk-israel-apa-efeknya#:~:text=BDS adalah gerakan boikot \(penolakan,memberikan hak setara kepada Palestina.](https://www.cnbcindonesia.com/news/20231114170902-4-488970/ri-banyak-negara-boikot-produk-israel-apa-efeknya#:~:text=BDS adalah gerakan boikot (penolakan,memberikan hak setara kepada Palestina.)
- Supriyanto, J., Alita, D., & Isnain, A. R. (2023). Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 4(1), 74–80. <https://doi.org/10.33365/jatika.v4i1.2468>
- Wiradi, G. (2022). *Metode Penelitian Studi Literatur, Apa Itu?* 1. <https://buku.kompas.com/read/2051/metode%EF%BF%BEpenelitian-studi-literatur-apa-itu>