

Public Sentiment Analysis on the Boycott Israel Movement on Platform X Using Random Forest and Logistic Regression Algorithms

Rachmayanti Tri Agustin*, Yana Cahyana^{2**}, Kiki Ahmad Baihaqi^{3*}, Tatang Rohana^{4**}

* Teknik Informatika, Universitas Buana Perjuangan Karawang

if21.rachmayantiagustin@mhs.ubpkarawang.ac.id¹, yana.cahyana@ubpkarawang.ac.id², kikiahmad@ubpkarawang.ac.id³, tatang.rohana@ubpkarawang.ac.id⁴

Article Info

Article history:

Received 2025-04-29

Revised 2025-06-03

Accepted 2025-06-05

Keyword:

Boycott,
Logistic Regression,
Random Forest,
Sentiment Analysis,
Social Media.

ABSTRACT

This research aims to analyze public sentiment toward the boycott movement against Israel on the X platform by applying Random Forest and Logistic Regression algorithms. The study uses 616 tweets collected through web crawling with relevant keywords such as "Boikot", "Israel", and "Palestine", covering the period from March 1, 2023 to January 30, 2025. The dataset underwent preprocessing including cleaning, normalization, stopword removal, tokenization, and stemming. Sentiment labeling was conducted both manually, categorizing the data into positive, negative, and neutral classes. TF-IDF was used for feature weighting. The data was split into 80% training and 20% testing. The Random Forest model achieved an accuracy of 70%, while Logistic Regression reached 68%. Both models showed higher accuracy in predicting positive sentiment compared to negative and neutral. The results suggest that public opinion on the boycott movement on social media tends to be supportive, with "Boikot," "Israel," and "Palestine" being the most dominant terms. Random Forest performed slightly better in classification, though improvements are needed in recognizing non-positive sentiments.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pada era digital saat ini, platform X telah menjadi media sosial yang sangat digemari banyak orang di berbagai kalangan yang memungkinkan pengguna untuk menyampaikan opini, perasaan, dan pandangan secara luas. X menjadi salah satu media sosial yang ramai digunakan oleh jutaan pengguna setiap harinya untuk berbagi pemikiran dan perasaan mereka terkait berbagai topik [1]. Gerakan boikot terhadap israel telah menjadi topik yang semakin banyak diperbincangkan, di Tengah memuncaknya pertentangan antara Palestina dan Israel [2], masyarakat Indonesia memberikan respon besar melalui platform x, yang turut menjadi bagian dari gerakan dukungan global yang terus meluas [3]. Aksi boikot terhadap produk yang berkaitan dengan Israel menunjukan solidaritas yang kini berkembang, bukan lagi sekedar urusan pribadi, melainkan sudah menjadi fenomena social yang meluas di berbagai media social seperti X, Tiktok, dan Instagram[4]. Hal ini menjadikan

analisis sentimen publik terhadap gerakan boikot Israel di media sosial sebagai topik yang relevan untuk dikaji.

Meskipun gerakan boikot Israel telah banyak dibahas di berbagai media, masih terdapat kesenjangan dalam memahami bagaimana sentimen publik terbentuk dan berkembang di media sosial. Kondisi saat ini menunjukkan bahwa banyak opini publik yang muncul di media sosial bersifat beragam dan sering kali tidak terstruktur [5]. Analisis sentimen memberikan wawasan yang mendalam tentang pandangan masyarakat terhadap gerakan tersebut. Penelitian ini menggunakan model *Random Forest*, yang merupakan kombinasi dari sejumlah pohon keputusan yang digabungkan menjadi satu model. Model ini tidak terpengaruh oleh waktu, sehingga hasil perhitungannya tetap konsisten kapanpun penelitian dilakukan. Selain itu, penelitian ini juga memanfaatkan *Logistic Regression*, sebuah teknik regresi yang digunakan untuk mengelompokkan dataset menjadi dua kategori. Namun, metode yang diterapkan adalah *Multinomial Logistic Regression*, yang memungkinkan

pengelompokan dataset menjadi 3 kelas, yaitu negatif, positif, dan netral [6]

Penelitian ini didasarkan pada studi sebelumnya yang memiliki keterkaitan dengan topik yang dibahas. (Agustina & Hendry, 2021) Dalam jurnal penelitian yang berjudul “Analisis Sentimen Terhadap Kontroversi Fatwa MUI Nomor 83 Tahun 2023 Tentang Pemboikotan Produk yang Terafiliasi Israel” penelitian ini menggunakan metode *Naïve Bayes*, *Decision Tree*, *Random Forest*, *Support Vector Machine (SVM)*, dan *K- Nearest Neighbor (KNN)*, hasil analisis sentimen tersebut menunjukkan 3 klasifikasi yaitu setuju, tidak setuju, dan netral dengan tingkat akurasi Naive Bayes 75%, Decision Tree 65%, Random Forest 67%, SVM 63%, K-NN 53% [7].

Tujuan dari penelitian ini yaitu untuk menganalisis pandangan masyarakat terkait boikot Israel melalui analisis sentimen pada social media X menggunakan model *Random Forest* dan *Logistic Regression*. Dengan menjadikan data dari aplikasi X sebagai media informasi, penelitian ini bertujuan untuk mengidentifikasi pola, tren, dan perbedaan sentimen yang terjadi pada kalangan pengguna Twitter terkait isu boikot tersebut.

II. METODE

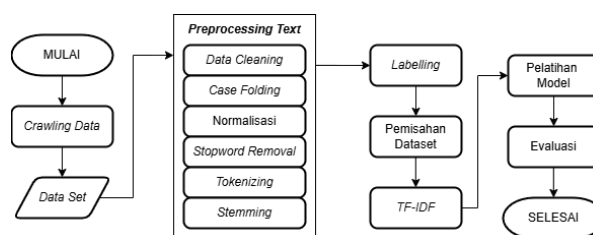
Pada tahap ini proses penelitian menggunakan algoritma *Random Forest* dan *Logistic Regression*, dalam proses klasifikasi terhadap analisis sentimen publik terhadap gerakan boikot israel pada platform X menggunakan algoritma *Random Forest* dan *Logistic Regression*, penggunaan algoritma ini dikarenakan algoritma Random Forest mempunyai kemampuan dalam menangani data yang kompleks serta ketahanannya terhadap overfitting. Algoritma ini merupakan ensemble dari beberapa decision tree yang bekerja secara kolektif untuk meningkatkan akurasi prediksi. Random Forest sangat efektif dalam menangani data teks yang telah diekstraksi menjadi fitur numerik, seperti TF-IDF, dan mampu mengidentifikasi pola kompleks dalam data [8]. Sementara itu, Logistic Regression dipilih sebagai pembanding karena merupakan algoritma yang sederhana, cepat, dan sangat cocok untuk klasifikasi biner, seperti sentimen positif atau negative. Meskipun tergolong model linear, logistic regression seringkali memberikan performa yang kompetitif dalam analisis sentimen, terutama pada dataset yang tidak terlalu kompleks [9]. Penggunaan dua algoritma ini memungkinkan evaluasi performa yang lebih menyeluruh serta membantu menentukan pendekatan mana yang lebih optimal untuk kasus analisis sentimen terhadap gerakan boikot israel di platform X. Proses ini melewati 9 tahapan penelitian, dimulai dari *Crawling Data*, *Preprocessing Text*, *Labelling*, *Pemisahan Dataset*, *TF-IDF*, *Pelatihan Model*, dan *Evaluasi*. Gambar 1 merupakan alur penelitian.

A. Crawling Data

Proses pengumpulan data dilakukan dari *platform X* menggunakan Teknik *Web Crawling* dengan memanfaatkan

auth token dari akun X, Proses *crawling* dilakukan untuk mengumpulkan *tweet-tweet* yang berisi opini atau sentimen terkait gerakan boikot Israel. Kata kunci seperti Boikot, Israel, Palestine digunakan sebagai parameter pencarian untuk menghasilkan data yang sesuai dengan topik penelitian.

Proses *crawling* dimulai dari autentikasi menggunakan *auth token*, yang memungkinkan akses ke konten publik pada *platform X*. data yang dikumpulkan meliputi teks *tweet*, tanggal, dan waktu posting, serta informasi tambahan seperti jumlah *likes* dan *retweet* jika diperlukan. Rentang waktu pengumpulan data dibatasi pada periode tertentu yaitu dari tanggal 1 Maret 2023 sampai 30 Januari 2025 hal itu untuk memastikan relevansi data dengan isu yang sedang dibahas.



Gambar 1. Alur Penelitian

Data hasil crawling memperoleh 616 data lalu disimpan dalam bentuk format CSV, yang memudahkan analisis lebih lanjut. Dalam hal ini, perhatian khusus diberikan pada etika penelitian, termasuk menjaga kerahasiaan data pengguna dan menjamin bahwa data yang diperoleh dipakai hanya untuk kepentingan penelitian. Tahap preprocessing dilakukan menggunakan data dari proses crawling tersebut dan digunakan untuk menganalisis sentimen untuk mengidentifikasi opini publik terhadap gerakan boikot Israel menggunakan model *Random Forest* dan *Logistic Regression*.

B. Preprocessing Data

Proses *preprocessing* meliputi *data cleaning*, *case folding*, normalisasi, *stopword removal*, tokenisasi, dan *stemming*. *Data cleaning* membersihkan karakter yang tidak diperlukan, *case folding* mengganti semua huruf kapital menjadi huruf kecil, dan normalisasi mengganti kata tidak baku menjadi baku. *Stopword removal* menghilangkan kata-kata umum yang kurang relevan, tokenisasi memecah kalimat menjadi kata-kata, dan *stemming* mengubah kata ke bentuk dasarnya [10]. Setelah dilakukan *preprocessing data*, jumlah data yang tersisa yaitu 616 data. Tahapan ini bertujuan untuk menyederhanakan dan menstandarisasi data agar analisis teks lebih akurat.

C. Labelling

Proses pelabelan data dalam penelitian ini dilakukan secara manual dengan bantuan seorang ahli pakar bahasa Indonesia, yaitu Adelya Dariyah, M.Pd. Beliau berperan penting dalam mengklasifikasikan sentimen pada data tweet menjadi tiga kategori, yaitu sentimen positif, negatif, dan netral. Penilaian dilakukan dengan mempertimbangkan konteks bahasa, makna tersirat, serta penggunaan diksi dalam setiap tweet

yang telah dikumpulkan. Keterlibatan ahli bahasa ini bertujuan untuk meningkatkan validitas dan akurasi pelabelan data, sehingga model yang dibangun dapat melakukan prediksi sentimen dengan lebih baik dan sesuai dengan karakteristik bahasa Indonesia yang digunakan di media sosial.

D. Pemisahan Dataset

Proses pemisahan data dibagi menjadi 2 yaitu data latih dan data uji yang bertujuan untuk membagi data dari hasil *preprocessing* ke dalam perbandingan tertentu. Jumlah data yang diperoleh dari preprocessing ada 616 lalu dibagi dengan perbandingan jumlah pembagian data dalam penelitian ini yaitu 80%:20% terdiri dari 80% data train sebanyak 492 dan 20% data test sebanyak 124. Pembagian dengan perbandingan tersebut dikarenakan semakin banyak data yang dilatih, maka akan memperoleh hasil analisis lebih akurat.

E. Pembobotan TF-IDF

Metode pembobotan kata dalam dokumen atau sering disebut juga TF-IDF bertujuan untuk mengonversi teks ke format numerik dan meningkatkan akurasi analisis [11]. Perhitungan TF-IDF melibatkan Term Frequency atau TF, yang mengukur frekuensi kata yang muncul pada dokumen, serta Inverse Document Frequency atau IDF, yang memberi bobot lebih kecil pada kata umum di seluruh dokumen. Nilai IDF dihitung menggunakan logaritma berdasarkan jumlah total dokumen dan dokumen yang mengandung kata tertentu [12].

F. Pelatihan Model

Penelitian ini menggunakan model Random Forest dan Logistic Regression guna menganalisis sentimen publik terhadap gerakan boikot Israel di platform X. Random Forest membangun beberapa pohon keputusan untuk menghasilkan prediksi yang lebih akurat, sementara Logistic Regression memodelkan hubungan antara fitur dan label sentimen menggunakan fungsi logit. Fitur teks dikonversi ke bentuk numerik, seperti TF-IDF, sebelum digunakan sebagai input model. Model dilatih pada data latih dan di evaluasi dengan data uji menggunakan metrik akurasi, precision, recall, f1-score untuk memastikan kinerja prediksi yang optimal.

G. Confusion Matrix

Confusion matrix yaitu metode evaluasi model klasifikasi untuk membandingkan hasil prediksi dengan label asli dalam bentuk tabel. Terdapat empat elemen utama: True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). Dari matriks ini, dapat dihitung metrik evaluasi seperti accuracy, precision, recall, dan f1 score [13].

Pada riset ini, proses confusion matrix bertujuan untuk evaluasi model Random Forest dan Logistic Regression dalam analisis sentiment gerakan boikot israel di platform X. Dengan matriks ini, dapat diketahui letak kesalahan prediksi serta performa keseluruhan model. Berikut merupakan rumus dari confusion matrix :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

Tahap ini akan menguraikan serta menjelaskan hasil penelitian yang menggunakan model *Random Forest* dan *Logistic Regression* untuk sentimen analisis, dengan pemrosesan dataset yang dilakukan menggunakan *Google Collab*.

A. Pengumpulan Data

Pada tahap pengumpulan data, metode *Crawling Data* diterapkan untuk mengumpulkan *tweet* menggunakan *TweetHarvest* di *Google Colab*. Proses crawling data dilakukan dengan menentukan kata kunci Boikot, Israel, dan Palestine sesuai dengan topik penelitian, setelah proses *crawling* selesai, diperoleh sebanyak 616 *tweet*, kemudian data disimpan dalam format CSV untuk keperluan analisis lebih lanjut. Berikut merupakan hasil crawling data menggunakan tweetharvest :

	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url
0	1879774384892206882	Thu Jan 16 12:51:12 +0000 2025	0	@tanyariffes Tetap boikot ya teman- teman. Boiko...	1879874042284609728	NaN
1	1879370138287489053	Wed Jan 15 03:28:52 +0000 2025	67	Mnggu lalu saat lat foto Kumaila Coki ... Lng...	1879370138287489053	NaN
2	187828147096906489	Tue Jan 14 16:28:48 +0000 2025	0	@Viannajasso @SatuBinu3299 @thinkub...lu nam...	1879203524019662998	NaN
3	1877626155387261859	Sat Jan 11 19:38:48 +0000 2025	1	Inget... kita juga kara sokong boikot yang tela...	1878164878314213693	https://pbs.twimg.com/media/GhCU0kMbiAARika.jpg

Gambar 2. Dataset

B. Preprocessing Data

Proses *preprocessing data* bertujuan agar mengoptimalkan kualitas data sebelum dilakukan analisis sentimen. Tahapan ini mencakup, *Data Cleaning* untuk menghapus karakter khusus, tanda baca, dan data duplikat, *Case Folding* yang mengonversi seluruh teks menjadi huruf kecil, serta Normalisasi untuk menyeragamkan kata tidak baku. Selain itu, dilakukan *Stopword Removal* untuk menghilangkan kata yang tidak memiliki nilai, *Tokenizing* untuk memecah teks menjadi kata-kata individu, *Stemming* guna mengubah kata ke bentuk dasarnya. Berikut merupakan tahapan preprocessing text :

1. Data Cleaning

Data Cleaning bertujuan untuk membersihkan data dari karakter khusus, tanda baca, angka, dan duplikasi yang tidak diperlukan dalam analisis sentimen [14]. Proses ini membantu meningkatkan kualitas data agar terlihat lebih terstruktur dan

relevan. Berikut merupakan tabel hasil dari sebelum data cleaning dan setelah data cleaning :

TABEL 1.
DATA CLEANING

Sebelum	Setelah
Ekonomi Israel makin menipis makin kita boikot produknya maka abis hartanya gabisa beli senjata buat nyerang Palestine lagi #AyoBoikotIsrael https://t.co/Daf57thHNv	Ekonomi Israel makin menipis makin kita boikot produknya maka abis hartanya gabisa beli senjata buat nyerang Palestine lagi

2. Case Folding

Proses *Case Folding* bertujuan untuk mengubah seluruh teks menjadi huruf kecil agar mempermudah proses analisis [15]. Dengan tahap ini, perbedaan huruf besar dan kecil tidak akan memengaruhi hasil pemrosesan data. Berikut merupakan tabel hasil dari proses case folding :

TABEL 2.
CASE FOLDING

Sebelum	Setelah
Ekonomi Israel makin menipis makin kita boikot produknya maka abis hartanya gabisa beli senjata buat nyerang Palestine lagi	ekonomi israel makin menipis makin kita boikot produknya maka abis hartanya gabisa beli senjata buat nyerang palestine lagi

3. Normalisasi

Normalisasi bertujuan untuk mengganti kata tidak baku dengan padanan yang sesuai aturan kebahasaan, sehingga lebih mudah dipahami oleh model analisis. Proses ini dilakukan untuk menyamakan berbagai bentuk penulisan dalam teks agar lebih konsisten. Berikut merupakan tabel hasil dari sebelum normalisasi dan setelah normalisasi :

TABEL 3.
NORMALISASI

Sebelum	Setelah
ekonomi israel makin menipis makin kita boikot produknya maka abis hartanya gabisa beli senjata buat nyerang palestine lagi	ekonomi israel makin menipis makin kita boikot produknya maka habis hartanya tidak bisa beli senjata untuk nyerang palestine lagi

4. Stopword Removal

Setelah tahap normalisasi, tahap berikutnya adalah *stopword removal* yang berguna untuk membuang kata umum yang tidak bermakna penting [16], seperti “dan”, “yang”, “di”, serta kata lainnya yang tidak memengaruhi konteks utama. Proses ini membantu mengurangi noise dalam data dan meningkatkan efisiensi analisis. Berikut merupakan tabel hasil dari sebelum *stopword removal* dan setelah *stopword removal* :

TABEL 4.
STOPWORD REMOVAL

Sebelum	Setelah
ekonomi israel makin menipis makin kita boikot produknya maka habis hartanya tidak bisa beli senjata untuk nyerang palestine lagi	ekonomi israel makin menipis makin kita boikot produknya maka habis hartanya tidak bisa beli senjata untuk nyerang palestine

5. Tokenizing

Tahap *Tokenizing* bertujuan untuk membagi teks menjadi kata-kata atau token yang lebih kecil [17] agar dapat diproses lebih lanjut dalam analisis sentiment. Berikut merupakan tabel hasil dari sebelum *tokenizing* dan setelah *tokenizing* :

TABEL 5.
TOKENIZING

Sebelum	Setelah
ekonomi israel makin menipis makin kita boikot produknya maka habis hartanya tidak bisa beli senjata untuk nyerang palestine	['ekonomi', 'israel', 'makin', 'menipis', 'makin', 'kita', 'boikot', 'produknya', 'maka', 'habis', 'hartanya', 'tidak', 'bisa', 'beli', 'senjata', 'untuk', 'palestine']

6. Stemming

Proses *Stemming* berguna untuk mengembalikan kata berimbuhan menjadi bentuk dasarnya agar lebih mudah diproses dalam analisis sentiment [18]. Berikut merupakan tabel hasil dari sebelum *stemming* dan setelah *stemming* :

TABEL 6.
STEMMING

Sebelum	Setelah
['ekonomi', 'israel', 'makin', 'menipis', 'makin', 'kita', 'boikot', 'produknya', 'maka', 'habis', 'hartanya', 'tidak', 'bisa', 'beli', 'senjata', 'untuk', 'nyerang', 'palestine']	['ekonomi', 'israel', 'makin', 'menipis', 'makin', 'kita', 'boikot', 'produknya', 'maka', 'habis', 'hartanya', 'tidak', 'bisa', 'beli', 'senjata', 'untuk', 'nyerang', 'palestine']

C. Labelling

Pelabelan data dilakukan secara manual oleh pakar Bahasa Indonesia yaitu Adelya Dariyah, M.Pd untuk menjamin ketepatan klasifikasi setiap data. Dengan keahlian linguistik yang dimiliki, proses labelling berjalan secara sistematis dan sesuai konteks Bahasa. Hasil dari pelabelan ini menjadi acuan utama dalam proses pelatihan model, sehingga membantu meningkatkan akurasi dan keandalan sistem dalam mengenali pola Bahasa secara optimal. Berikut merupakan hasil labelling menggunakan bantuan pakar Bahasa Indonesia :

TABEL 7.
DATASET LABEL

No.	Text	Label
1.	kita boikot selebriti jadi ba aktif diendorse awal serang israel	Positive

	palestine di bulan okt 2023 detik kalo bilang harus di boikot produk aja jangan artis aku rasa udah masuk nyata yang nggak tepat	
2.	cape cape masyarakat indonesia teriak boikot produk pro israel eh cawapresnya malah kek gin gara gara satu foto ngerusak nama baik indonesia ranah internasional negara negara pro Palestine	Negative
3.	oalah nyata cerita pki sulit lah amerika main lawan amerika ga kalau udah negara turut campur kurang pengaruh cina rusia saat komunis asia tenggara	Neutral

Tabel berikut menunjukkan distribusi jumlah tweet berdasarkan label sentiment.

TABEL 8.
TWEET PERKELAS

Sentimen	Jumlah Tweet
Positive	406
Negative	173
Neutral	37
Total	616

Berdasarkan distribusi data, terdapat 406 tweet berlabel positive, 173 tweet berlabel negative, dan hanya 37 tweet berlabel neutral dari total 616 data. Hal ini menunjukkan bahwa distribusi kelas bersifat tidak seimbang, dengan dominasi kelas positif. Ketidakseimbangan ini dapat memengaruhi performa model klasifikasi, terutama dalam mengenali kelas yang jumlahnya lebih sedikit seperti neutral.

Sementara hasil observasi terhadap keseluruhan dataset, mayoritas tweet ditulis dalam Bahasa Indonesia, yang menunjukkan bahwa partisipasi diskusi mengenai Gerakan boikot Israel di platform X Sebagian besar berasal dari wilayah Indonesia. Meskipun demikian, terdapat pula sejumlah frasa atau kutipan dalam Bahasa Inggris, khususnya pada bagian retweet atau kutipan berita internasional, yang mengindikasikan adanya pengaruh dari wacana global dalam percakapan tersebut.

D. Pemisahan Dataset

Pembagian data pada penelitian ini menggunakan proporsi 80:20. Di mana 80% digunakan sebagai data latih yang diperoleh sebanyak 492 dan 20% sebagai data uji yang diperoleh sebanyak 124. Pembagian ini bertujuan untuk memastikan model dapat belajar secara optimal sekaligus diuji dengan data yang representative.

E. Pembobotan TF-IDF

Pembobotan TF-IDF dalam penelitian ini dilakukan untuk mengetahui kata-kata yang memiliki pengaruh besar dalam dokumen. Hasil tertinggi menunjukkan kata “Boikot” dengan skor 6.86, diikuti oleh “Israel” dengan skor 6.83, dan “Palestine” dengan skor 6.00. Kata-kata tersebut memiliki

frekuensi tinggi dan bobot penting dalam konteks data yang di analisis. Gambar 4 menunjukkan hasil dari TF-IDF :

	Kata	TF-IDF	Frekuensi
183	boikot	6.861798	141
516	israel	6.837178	134
938	palestine	6.009934	112
377	ga	4.053718	39
332	dukung	3.474062	31

Gambar 3. TF-IDF

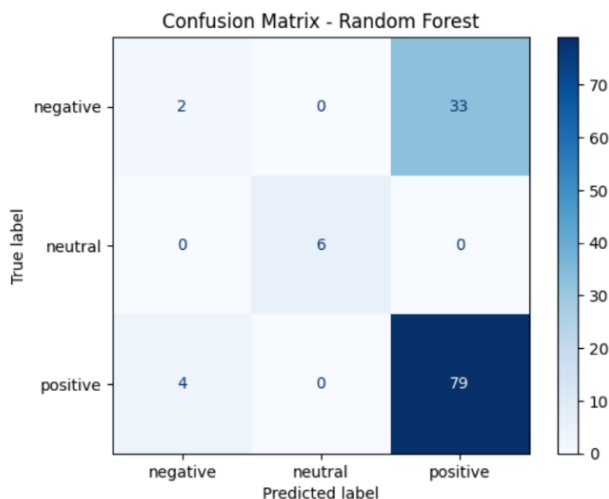
F. Random Forest

Dari hasil evaluasi model Random Forest yang ditunjukkan pada Classification Report, diketahui bahwa model menghasilkan nilai akurasi sebesar 0,70. Hasil tersebut menunjukkan bahwa model cukup baik dalam melakukan klasifikasi, meskipun masih terdapat ketidakseimbangan performa antar kelas. Model cenderung lebih akurat dalam memprediksi kelas positive, sementara performa untuk kelas negative dan neutral masih rendah. Berikut merupakan hasil dari klasifikasi menggunakan model random forest :

=== Classification Report ===				
	precision	recall	f1-score	support
negative	0.33	0.06	0.10	35
neutral	1.00	1.00	1.00	6
positive	0.71	0.95	0.81	83
accuracy			0.70	124
macro avg	0.68	0.67	0.64	124
weighted avg	0.61	0.70	0.62	124

Gambar 4. Evaluasi Random Forest

Evaluasi terhadap model Random Forest dilakukan menggunakan data uji, dengan hasil akurasi sebesar 0.70, precision sebesar 0.61, recall sebesar 0.70, dan F1-score sebesar 0,61. Model ini menunjukkan kemampuan klasifikasi yang cukup optimal berdasarkan hasil yang diperoleh, namun masih terdapat ruang untuk perbaikan. Berdasarkan confusion matrix pada gambar 5, terlihat bahwa model cenderung lebih baik dalam mengklasifikasikan data ke dalam kelas positif, sementara prediksi terhadap kelas negative dan neutral masih kurang akurat. Berikut merupakan hasil evaluasi menggunakan model random forest :



Gambar 5. Confusion Matrix Random Forest

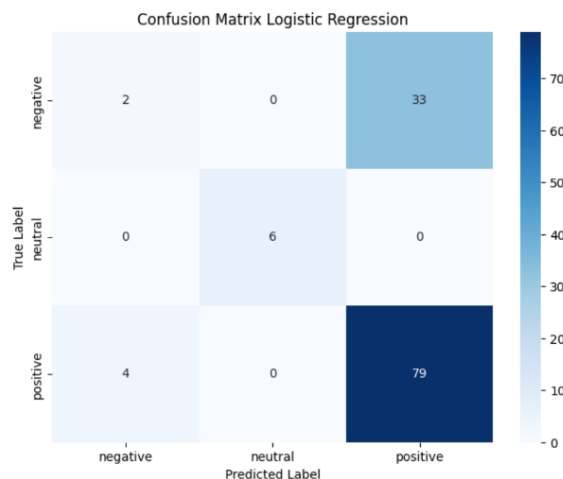
G. Logistic Regression

Hasil evaluasi dari *Logistic Regression* yang digunakan dalam riset ini menunjukkan hasil akurasi sebesar 0,68. Hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang cukup, meskipun belum optimal. Model cenderung hanya mampu mengenali kelas positive dengan baik, sedangkan performa untuk kelas negative dan neutral masih rendah. Berikut merupakan hasil dari klasifikasi menggunakan model logistic regression :

Classification Report:				
	precision	recall	f1-score	support
negative	0.57	0.11	0.19	35
neutral	1.00	0.00	0.00	6
positive	0.68	0.96	0.80	83
accuracy			0.68	124
macro avg	0.75	0.36	0.33	124
weighted avg	0.67	0.68	0.59	124

Gambar 6. Evaluasi Logistic Regression

Evaluasi terhadap model Logistic Regression dilakukan menggunakan data uji, dengan hasil akurasi sebesar 0,68. Nilai precision yang diperoleh sebesar 0,61, recall sebesar 0,67, dan F1-score sebesar 0,58. Hasil ini menunjukkan bahwa performa model tergolong cukup baik, meskipun masih memerlukan perbaikan agar hasil klasifikasi lebih merata di semua kelas. Berdasarkan confusion matrix pada Gambar 7, terlihat bahwa model sangat dominan dalam memprediksi kelas positive, namun memiliki kelemahan dalam mengklasifikasikan kelas negative dan neutral. Berikut merupakan hasil evaluasi menggunakan model logistic regression:



Gambar 7. Confusion Matrix Logistic Regression

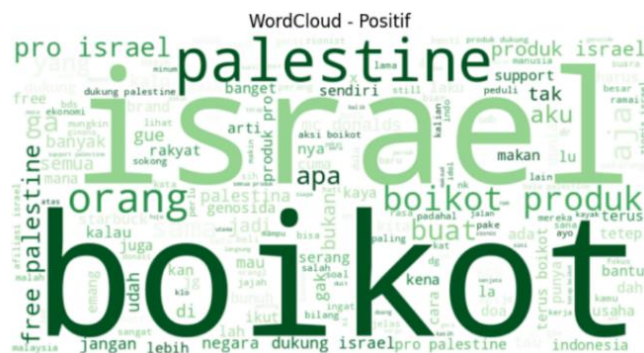
Berikut merupakan hasil perbandingan dari evaluasi model Random Forest dan Logistic Regression :

TABEL 9.
PERBANDINGAN MODEL

Model	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
<i>Random Forest</i>	0,70	0,61	0,70	0,61
<i>Logistic Regression</i>	0,68	0,61	0,67	0,58

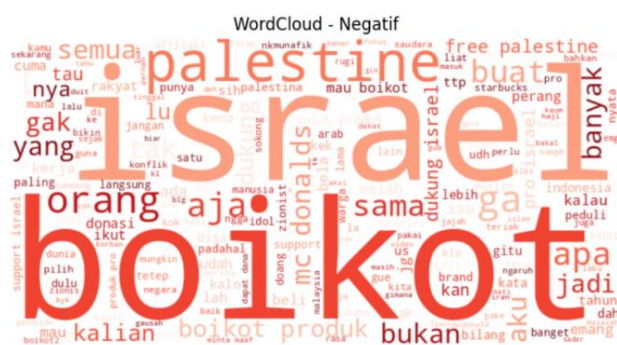
H. Visualisasi

Visualisasi WordCloud untuk sentimen positif yang ditunjukkan pada Gambar 10, kata-kata yang paling dominan adalah "*boikot*", "*israel*", "*palestine*", dan "*produk*". Hal ini menunjukkan bahwa dalam kumpulan data tersebut, opini positif banyak terkait dengan dukungan terhadap gerakan boikot terhadap produk Israel sebagai bentuk solidaritas terhadap Palestina. Kata-kata seperti "*dukung*", "*rakyat*", "*free palestine*", dan "*genosida*" juga muncul dengan frekuensi tinggi, yang mengindikasikan bahwa sentimen positif dalam konteks ini lebih mengarah pada ekspresi dukungan kemanusiaan dan keadilan sosial terhadap isu konflik Palestina-Israel.



Gambar 8. Wordcloud Positif

Visualisasi *WordCloud* negatif pada Gambar 11, terlihat bahwa kata-kata yang paling menonjol adalah "boikot", "israel", "palestine", "orang", dan "kalian". Hal ini menunjukkan bahwa percakapan negatif yang dianalisis dalam konteks ini banyak berkaitan dengan isu boikot terhadap produk atau afiliasi yang berhubungan dengan Israel, serta dukungan terhadap Palestina. Kata-kata lain seperti "mc donalds", "starbucks", "donasi", "dukungan", dan "konflik" juga muncul cukup sering, mencerminkan sentimen publik yang kuat terhadap isu-isu kemanusiaan dan politik yang sedang berlangsung. Dominasi kata-kata tersebut mengindikasikan bahwa percakapan negatif didorong oleh respons emosional terhadap konflik geopolitik, serta seruan aksi kolektif dari masyarakat dalam bentuk boikot.



Gambar 9. Wordcloud Negatif

Selain analisis kuantitatif dengan model Random Forest dan Logistic Regression, pemahaman konteks geopolitik juga krusial dalam melihat fluktuasi sentimen publik terhadap gerakan boikot Israel. Lonjakan sentimen negatif tercatat pada awal Oktober 2023, bertepatan dengan eskalasi konflik Israel-Palestina, khususnya serangan intensif ke Gaza yang memicu gelombang kemarahan publik di media sosial.

Analisis isi tweet menunjukkan bahwa ajakan boikot dari tokoh publik dan brand tertentu, seperti selebriti yang mendukung Palestina, turut mendorong tingginya interaksi retweet, likes, dan komentar pada tweet bertema boikot.

Respons publik pun beragam: sebagian besar mendukung boikot sebagai bentuk solidaritas, sementara sebagian lain mempertanyakan efektivitas atau konsistensi aksi tersebut. Hal ini menegaskan bahwa opini publik dibentuk oleh emosi, konteks geopolitik, dan pengaruh tokoh yang menyuarakan aksi boikot [19].

IV. KESIMPULAN

Penelitian ini menganalisis sentimen public terhadap Gerakan boikot Israel di platform X menggunakan algoritma Random Forest dan Logistic Regression. Tahap klasifikasi pada penelitian ini dilakukan dengan menggunakan data latih. Mengacu pada hasil evaluasi model Random Forest yang ditunjukkan pada classification report, diketahui bahwa model menghasilkan nilai akurasi sebesar 0,70, sedangkan model Logistic Regression menghasilkan akurasi sebesar 0,68.

Secara keseluruhan, model Random Forest menunjukan performa yang sedikit lebih baik dibandingkan Logistic

Regression. Kedua model memiliki kecenderungan untuk lebih akurat dalam mengklasifikasikan sentimen positif, namun masih kurang optimal dalam mengenali sentimen negative dan netral.

Model ringan seperti Logistic Regression dan SVM memiliki keunggulan dalam interpretabilitas serta kemudahan untuk di-deploy secara cepat di lingkungan produksi dengan sumber daya terbatas. Meskipun model seperti XGBoost menawarkan performa yang lebih tinggi, pertimbangan efisiensi dan kebutuhan aplikasi real-time membuat model sederhana tetap relevan dalam banyak scenario.

DAFTAR PUSTAKA

- [1] J. Khatib Sulaiman, D. Atmajaya, A. Febrianti, H. Darwis, I. Artikel Abstrak, And K. Kunci, "Metode Svm Dan Naive Bayes Untuk Analisis Sentimen Chatgpt Di Twitter," *Indonesian Journal Of Computer Science Attribution*, Vol. 12, No. 4, P. 2173, Doi: <https://doi.org/10.33022/Ijcs.V12i4.3341>.
- [2] Isalman, Ilyas, Farhan Ramadhani Istianandar, And Nurul Ittaquillah, "Boycott Campaign Intensity On Consumer Boycott Intentions And Participation: The Role Of Access To Substitute Products," *Journal Of Economics, Business, And Accountancy Ventura*, Vol. 27, No. 3, Pp. 430–444, Mar. 2025, Doi: [10.14414/Jebav.V27i3.4737](https://doi.org/10.14414/Jebav.V27i3.4737).
- [3] S. Fatimah Azzahro, "Sentiment Analysis On Boycott Movement: Nvivo Approach," 2024. [Online]. Available: <https://ejournal.stiesyariah bengkalis.ac.id/index.php/Iqtishadun>
- [4] P. Wahyuni And M. A. Romli, "Comparison Of Naïve Bayes Classifier And Decision Tree Algorithms For Sentiment Analysis On The House Of Representatives' Right Of Inquiry On Twitter," 2024. Doi: <https://doi.org/10.30871/Jaic.V8i2.8670>.
- [5] Ade Tiara Susilawati, Nur Anjeni Lestari, And Puput Alpria Nina, "Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naïve Bayes," *Nian Tana Sikka : Jurnal Ilmiah Mahasiswa*, Vol. 2, No. 1, Pp. 26–35, Dec. 2023, Doi: [10.59603/Niantanasikka.V2i1.240](https://doi.org/10.59603/Niantanasikka.V2i1.240).
- [6] N. Agung, P. #1, And H. Bunyamin, "Perbandingan Logistic Regression Dengan Random Forest Dalam Memprediksi Sentimen Pada Imdb Movie Review,"
- [7] M. Yasir, M. Grace Haque, R. Suraji, And C. Author, "Analisis Sentimen Terhadap Kontroversi Fatwa Mui Nomor 83 Tahun 2023 Tentang Pemboikotan Produk Yang Terafiliasi Israel", Doi: [10.31933/Jemsi.V5i4](https://doi.org/10.31933/Jemsi.V5i4).
- [8] M. Rusdi Rahman, A. Febri Diansyah, And H. Hanafi, "Sentiment Analysis On The Shopee Application On Playstore Using The Random Forest Classification Method," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, Vol. 9, No. 1, Pp. 20–24, Nov. 2023, Doi: [10.25139/Inform.V9i1.5465](https://doi.org/10.25139/Inform.V9i1.5465).
- [9] Alisya Mutia Mantika, Agung Triayudi, And Rima Tamara Aldisa, "Sentiment Analysis On Twitter Using Naïve Bayes And Logistic Regression For The 2024 Presidential Election," *Sana: Journal Of Blockchain, Nfts And Metaverse Technology*, Vol. 2, No. 1, Pp. 44–55, Feb. 2024, Doi: [10.58905/Sana.V2i1.267](https://doi.org/10.58905/Sana.V2i1.267).
- [10] F. Mulya, S. Putra, S. Rakasiwi, And N. Ariyanto, "Twitter Sentiment Classification Towards Telecommunication Provider Users In Indonesia," 2025. Doi: <https://doi.org/10.30871/Jaic.V9i2.9143>.
- [11] I. P. Rahayu, A. Fauzi, And J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," *Jurnal Sistem Komputer Dan Informatika (Json)*, Vol. 4, No. 2, P. 296, Dec. 2022, Doi: [10.30865/Json.V4i2.5381](https://doi.org/10.30865/Json.V4i2.5381).
- [12] B. Ramadhani And R. R. Suryono, "Komparasi Algoritma Naïve Bayes Dan Logistic Regression Untuk Analisis Sentimen

- Metaverse,” *Jurnal Media Informatika Budidarma*, Vol. 8, No. 2, P. 714, Apr. 2024, Doi: 10.30865/Mib.V8i2.7458.
- [13] I. Habib Kusuma And N. Cahyono, “Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor,” Vol. 8, No. 3, 2023, Doi: <https://doi.org/10.30591/jpit.V8i3.5734>.
- [14] M. R. Nurhusen, J. Indra, And K. A. Baihaqi, “Analisis Sentimen Pengguna Twitter Terhadap Kenaikan Harga Bahan Bakar Minyak (Bbm) Menggunakan Metode Logistic Regression,” *Jurnal Media Informatika Budidarma*, Vol. 7, No. 1, P. 276, Jan. 2023, Doi: 10.30865/Mib.V7i1.5491.
- [15] R. A. Ramadhan, T. Rohana, T. Al Mudzakir, And D. Wahiddin, “Penerapan Algoritma Support Vector Machines Dan Random Forest Dalam Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital,” *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, Vol. 7, No. 2, P. 969, Dec. 2024, Doi: 10.37600/Tekinkom.V7i2.1595.
- [16] E. R. Lidinillah, T. Rohana, And A. R. Juwita, “Analisis Sentimen Twitter Terhadap Steam Menggunakan Algoritma Logistic Regression Dan Support Vector Machine,” *Teknosains : Jurnal Sains, Teknologi Dan Informatika*, Vol. 10, No. 2, Pp. 154–164, Jul. 2023, Doi: 10.37373/Tekno.V10i2.440.
- [17] A. M. Siregar, “Analisis Sentimen Pindah Ibu Kota Negara (Ikn) Baru Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine (Svm),” *Faktor Exacta*, Vol. 16, No. 3, Oct. 2023, Doi: 10.30998/Faktorexacta.V16i3.16703.
- [18] A. Rizky Gunawan, R. Faticha, And A. Aziza, “Sentiment Analysis Using Lstm Algorithm Regarding Grab Application Services In Indonesia,” 2025. Doi: <https://doi.org/10.30871/jaic.V9i2.8696>.
- [19] N. Salwa Amarta And I. Susila, “Analysis Of The Influence Of Religious Beliefs And Solidarity On Product Boycott Behavior,” 2025.