

# IMPLEMENTASI MODEL BERT PADA ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP AKSI BOIKOT PRODUK ISRAEL

Zilvi Azus Sriyanti<sup>1\*</sup>, Dhian Satria Yudha Kartika<sup>2</sup>, Abdul Rezha Efrat Najaf<sup>3</sup>

<sup>1,2,3</sup>UPN “Veteran” Jawa Timur; Jalan Rungkut Madya No.1, Gunung Anyar, Kec. Gunung Anyar, Surabaya, Jawa Timur 60294, Indonesia; (031) 8706369

Received: 1 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

## Keywords:

Analisis sentimen;

Twitter;

Boikot;

BERT.

## Correspondent Email:

rezha.efrat.sifo@upnjatim.ac.id

**Abstrak.** Konflik antara Palestina dan Israel telah terjadi selama beberapa dekade. Baru-baru ini konflik kembali memanas dan menimbulkan banyak korban warga sipil. Adanya konflik ini menimbulkan sebuah aksi yaitu aksi boikot produk Israel dengan harapan bisa menekan perekonomian Israel. Aksi boikot di Indonesia menjadi sorotan masyarakat dan menimbulkan tanggapan positif maupun negatif yang diungkapkan melalui *Twitter*. Penelitian ini memiliki tujuan untuk menganalisis sentimen terhadap aksi boikot produk Israel pada *Twitter* untuk memperoleh pandangan yang mendalam mengenai aksi boikot produk Israel. Penelitian ini menggunakan pelabelan otomatis dengan *Vader Lexicon* dan melakukan penyeimbangan kelas sebelum klasifikasi menggunakan *Random Oversampling*. Penelitian ini menggunakan model BERT dalam bahasa Indonesia untuk melakukan klasifikasi dan pengujian menggunakan *confusion matrix*. Nilai akurasi model yang didapatkan dalam penelitian ini yaitu 85%, presisi 85%, *recall* 85%, dan *f1-score* 85%. Visualisasi ditampilkan dalam *website* menggunakan *streamlit*. Sentimen yang banyak diungkapkan masyarakat yaitu sentimen positif yang mengarah pada dukungan untuk melakukan aksi boikot produk Israel. Hasil penelitian ini diharapkan dapat memberikan pemahaman respon publik terhadap aksi boikot produk Israel di Indonesia.

**Abstract.** The conflict between Palestine and Israel has been going on for decades. Recently, the conflict has heated up again and caused many civilian casualties. This conflict gave rise to an action, namely a boycott of Israeli products in the hope of suppressing the Israeli economy. The boycott action in Indonesia became the public spotlight and gave rise to positive and negative responses expressed via *Twitter*. This research goals to analyze sentiment towards the boycott of Israeli products on *Twitter* to gain an in depth view of the boycott of Israeli products. This research uses automatic labeling with *Vader Lexicon* and performs class balancing before classification using *Random Oversampling*. This research uses the BERT model in Indonesian to carry out classification and testing using a *confusion matrix*. The model accuracy values obtained in this research were 85%, precision 85%, recall 85%, and *f1-score* 85%. Visualization is displayed on the website using *streamlit*. The sentiment expressed by many people is positive sentiment which leads to support for boycotting Israeli products. It is hoped that the results of this research provide an understanding of the public response to the boycott of Israeli products in Indonesia.

## 1. PENDAHULUAN

Baru-baru ini, telah terjadi konflik bersenjata antara kelompok militan dari Palestina ( Hamas ) dan Israel. Pada tanggal 7 Oktober 2023, Hamas menembakkan sekitar 5.000 roket menuju 22 daerah di luar jalur Gaza [1]. Konflik ini telah menyebabkan kerugian yang signifikan seperti memakan banyak korban diantaranya perempuan, warga sipil, dan anak-anak [2]. Adanya konflik ini menimbulkan sebuah aksi untuk melakukan boikot produk Israel untuk menekan perekonomian Israel. Di Indonesia, adanya Fatwa MUI Nomor 83 Tahun 2023 telah mendorong aksi masyarakat untuk tidak melakukan transaksi dan tidak menggunakan produk afiliasi Israel untuk mendukung Palestina [3].

Kontroversi terkait aksi boikot produk Israel telah menjadi perbincangan hangat masyarakat yang diungkapkan melalui media sosial. Media sosial memberikan pengaruh besar dalam membentuk pandangan masyarakat terkait boikot produk Israel [4]. Media sosial yang terkenal di Indonesia salah satunya yaitu Twitter. Presentase pemakai *Twitter* di Indonesia yaitu sebesar 60,2% menurut data dari *We Are Social* dan *Hootsuite* tahun 2023. Melalui *Twitter*, masyarakat dapat memberikan suara dan opini mereka terhadap aksi boikot produk Israel.

Dalam penelitian ini, analisis sentimen dilakukan menggunakan model (*Bidirectional Encoder Representations from Transformers*) untuk klasifikasi sentimen. BERT tersedia untuk banyak bahasa, termasuk bahasa Indonesia yaitu IndoBERT yang telah dilatih sekitar lebih dari 220 juta kata dalam Bahasa Indonesia. Pemilihan model ini didasarkan pada penelitian Husin mengenai komparasi *Algoritma Random Forest*, *Naïve Bayes*, dan BERT untuk melakukan *Multi-Class Classification* pada artikel *Cable News Network (CNN)*. Pada penelitian tersebut, BERT memiliki akurasi yang paling baik yaitu sebesar 92% dan *marco avg* dari *f1 score* sebesar 92% [5].

Oleh karena itu, penelitian ini memiliki tujuan untuk menganalisis sentimen khususnya masyarakat Indonesia terkait aksi boikot produk Israel menggunakan BERT bahasa Indonesia dari data *Twitter*. Hasil penelitian ini diharapkan dapat mengetahui analisis sentimen

masyarakat dan mengetahui nilai akurasi dari BERT bahasa Indonesia dalam melakukan klasifikasi sentimen terhadap aksi boikot produk Israel di Indonesia. Dengan demikian, diharapkan dapat memberikan pemahaman respon publik terhadap aksi boikot produk Israel di Indonesia.

## 2. TINJAUAN PUSTAKA

### 2.1. Boikot

Boikot adalah strategi yang digunakan untuk sejarah perjuangan rakyat untuk menentang ketidakadilan. Boikot digunakan untuk melakukan protes dan penolakan terhadap kebijakan suatu negara. Tindakan boikot diharapkan dapat mempengaruhi kebijakan negara yang dituju dengan menekan perekonomian [6].

### 2.2. Analisis Sentimen

Analisis sentimen termasuk dalam bidang pemrosesan *natural language* yang bertujuan untuk mengamati dan menganalisis opini, sikap, dan emosi yang dinyatakan oleh individu baik secara tertulis maupun lisan terkait dengan topik, produk, layanan, organisasi, atau kegiatan tertentu untuk menilai seberapa individu tersebut setuju atau tidak setuju dengan topik yang sedang diperbincangkan [7] [8].

### 2.3. Vader

*Valence Aware Dictionary Sentiment Reasoner* (VADER) adalah model leksikal yang digunakan untuk menilai sentimen seperti dalam *Twitter*. Terdapat empat kategori utama yaitu *compound*, *pos*, *neg*, dan *neu* untuk memberikan skor sentimen. *Compound score* merupakan hasil penjumlahan dari skor valensi dari kata dalam leksikon lalu dilakukan normalisasi menjadi rentang -1 (sangat negatif) sampai +1 (sangat positif) [9].

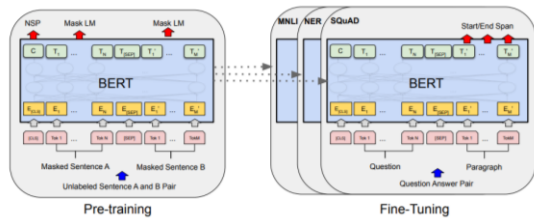
### 2.4. Random Oversampling

*Random Oversampling (ROS)* adalah proses yang digunakan untuk menambahkan data acak dari kelas minoritas ke dalam dataset. Proses menambahkan data acak dengan pengulangan hingga jumlah kelas dari kelas minoritas sama dengan jumlah kelas dalam kelas mayoritas. Data kelas minoritas dipilih *random* dan ditambahkan ke dataset agar jumlah data di

kelas minoritas ini seimbang dengan jumlah data kelas mayoritas [10].

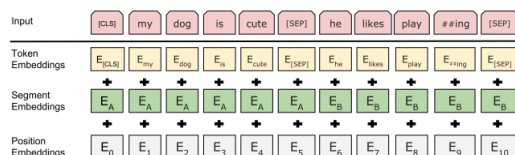
## 2.5. BERT

*Bidirectional Encoder Representations from Transformers* (BERT) adalah model bahasa yang diperkenalkan oleh Google tahun 2018. BERT mempunyai dua tahap kerja yaitu *pre-training* dan *fine-tuning*.



Gambar 1. Prosedur *Pre-training* dan *Fine-tuning* BERT[11]

Gambar 1 menjelaskan tahapan dalam *pre-training* dan *fine-tuning*. *Pre-training* melibatkan model dalam memahami bahasa dan konteksnya melalui *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP) bersamaan. MLM memungkinkan model menggabungkan konteks dari kedua arah, yaitu konteks kiri dan kanan. NSP menggabungkan representasi pasangan teks selama *pre-train* [11].



Gambar 2. Representasi *Input* BERT[11]

Gambar 2 menunjukkan representasi input dari BERT. Setiap urutan dimulai dengan token khusus [CLS]. Token khusus [SEP] ditambahkan untuk memisahkan setiap kalimat. Setiap token memiliki *segment embedding* yang membedakan antara kalimat A dan B. *Positional embedding* untuk memberikan tanda posisi setiap token dalam urutan. *Input encoder* BERT dihasilkan dengan menggabungkan *Token Embeddings*, *Segment Embeddings*, dan *Positional Embeddings*.

## 2.6. Confusion Matrix

*Confusion matrix* merupakan sebuah instrumen yang digunakan untuk mengevaluasi sistem klasifikasi yang menghasilkan tabel matriks untuk membandingkan hasil dari prediksi model dengan kelas sebenarnya dari

data [12]. Rumus *confusion matrix* ditunjukkan pada tabel 1.

Tabel 1. Rumus *confusion matrix*

Kelas		Kelas Prediksi	
		Kelas 1	Kelas 2
Kelas sebenarnya	Kelas 1	True Positive (TP)	False Negative (FN)
	Kelas 2	False Positive (FP)	True Negative (TN)

Keterangan :

1. *True Positive* (TP) = kondisi data yang diprediksi positif dan nilai sebenarnya positif.
2. *False Positive* (FP) = kondisi data yang diprediksi positif dan nilai sebenarnya negatif.
3. *False Negative* (FN) = kondisi data yang diprediksi negatif dan nilai sebenarnya positif.
4. *True Negative* (TN) = kondisi data yang diprediksi negatif dan nilai sebenarnya negatif.

Berikut merupakan rumus untuk menghitung *accuracy*, *precision*, *recall*, *f1-score*.

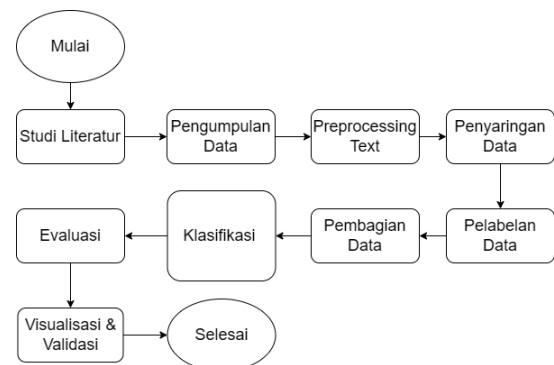
$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (1)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

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

## 3. METODE PENELITIAN



Gambar 3. Metodologi penelitian

Gambar 3 merupakan metodologi yang digunakan dalam penelitian. Berikut merupakan penjabaran dari setiap tahapan.

### 3.1. Studi Literatur

Tahapan ini melakukan pencarian rujukan teori terkait penelitian dari jurnal, buku, artikel ilmiah, penelitian terdahulu, dan sumber lain dari internet.

### 3.2. Pengumpulan Data

Tahapan ini melakukan pengumpulan data dari *Twitter* dari tanggal 8 Januari 2023 sampai 31 Desember 2023 berdasarkan 3 kata kunci pencarian yaitu “boikot produk israel”, “aksi boikot,” dan “boikot produk zionis.” Pengumpulan data dilakukan dengan menggunakan *tweet-harvest*, kemudian disimpan dalam bentuk *csv*.

### 3.3. Preprocessing Text

Pada tahapan ini melakukan pemrosesan teks agar bisa memenuhi standar dari model dan siap untuk dilakukan pelatihan. Tahapan praproses teks sebagai berikut.

1. *Case Folding*  
Tahap ini membuat seluruh karakter dalam teks menjadi format huruf kecil
2. *Data Cleaning*  
Tahap ini melakukan pembersihan karakter berulang, tautan, nama pengguna, angka, tagar, simbol, spasi berlebih, dan tanda baca.
3. *Normalization*  
Tahap ini melakukan pengubahan kata-kata yang tidak standar menjadi standar.
4. *Stopwords Removal*  
Tahap ini melakukan penghapusan kata-kata tidak bermakna signifikan seperti 'dari', 'dan', 'di', 'ke', 'yang'.
5. *Stemming*  
Tahap ini menghapus afiks seperti sufiks, prefiks, dan konfiks.

### 3.4. Penyaringan Data

Pada tahap ini melakukan penyaringan *tweet* yaitu menghilangkan *tweet* selain Bahasa Indonesia. *Tweet* akan disaring untuk menghindari bahasa asing, bahasa daerah, serta bahasa melayu. Penyaringan data juga dilakukan untuk menghapus data duplikat.

### 3.5. Pelabelan Data

Pada tahap ini dilakukan pelabelan otomatis menggunakan *Vader Lexicon*. Pelabelan

otomatid dilakukan untuk menentukan label teks dalam kategori positif atau negatif.

### 3.6. Pembagian Data

Data akan dibagi menjadi dua yang terdiri dari data latih (*train*) dan data uji (*test*). Data akan dibagi dengan pembagian 80:20 yaitu 80% untuk data latih dan 20% untuk data uji.

### 3.7. Klasifikasi

Tahap klasifikasi akan dilakukan dengan menggunakan IndoBERT ('indobenchmark/indobert-base-p1') yang merupakan model BERT untuk bahasa Indonesia. Sebelum dilakukan pelatihan, dilakukan *word embedding* untuk mengonversi kata-kata atau token dalam sebuah kalimat menjadi vektor numerik yang dapat dimengerti oleh BERT. Tahapan *word embedding* yaitu sebagai berikut.

1. *Tokenisasi*  
Tokenisasi dilakukan untuk memisahkan teks menjadi unit kecil yaitu token.
2. *Menambahkan Token [CLS] dan [SEP]*  
Token [CLS] digunakan sebagai token klasifikasi pada awal teks, sedangkan token [SEP] digunakan sebagai token penanda akhir teks atau pemisah antar teks.
3. *Mengonversi Token ke ID*  
Token akan diubah menjadi ID sesuai dengan kamus model yang digunakan yaitu "indobenchmark/indobert-base-p1".
4. *Menambahkan Padding*  
Penambahan token padding dilakukan agar semua input teks memiliki panjang yang sama sesuai panjang maksimum yang ditentukan.
5. *Menambahkan Attention Mask*  
*Attention mask* digunakan untuk membedakan token yang sebenarnya dari token *padding*.
6. *Menambahkan Positional Embedding*  
Penambahan *positional embedding* untuk memberikan informasi posisi absolut dalam teks.

### 3.8. Evaluasi

Tahap evaluasi dilakukan untuk menilai seberapa baik model melakukan klasifikasi sentimen. Performa model akan dievaluasi menggunakan *confusion matrix*. Selanjutnya dilakukan perhitungan perhitungan *accuracy*, *precision*, *recall*, dan *F1-Score*.

### 3.9. Visualisasi dan Validasi

Tahap visualisasi melakukan penulisan kode untuk tampilan antarmuka agar mempermudah pengguna memahami hasil klasifikasi. Visualisasi dilakukan menggunakan *framework streamlit* dengan menampilkan grafik hasil sentimen, visualisasi klasifikasi teks dan *file csv*. Tahap validasi yaitu mencocokkan hasil prediksi dengan analisis sebenarnya.

## 4. HASIL DAN PEMBAHASAN

Berikut ini hasil dan pembahasan klasifikasi sentimen menggunakan model BERT yaitu sebagai berikut.

### 4.1. Pengumpulan Data

*Scrapping tweet* dilakukan dengan menggunakan *tweet-harvest*. Hasil *scrapping* yang telah dilakukan dengan 3 kata kunci mendapatkan sebanyak 11.483 *tweet*. Contoh hasil dari pengumpulan data pada tabel 2.

Tabel 2. Contoh hasil *scrapping*

Created_at	Username	Full_text
Mon Oct 09 15:36:55 +0000 2023	misterqy77	@TxdariHI Boikot produk Israel dan sekutunya, nah mereka gak bakal bisa bergerak. Biasanya kalau lagi war mereka lupa base nya sendiri sudah aman/belum
Mon Dec 04 15:45:17 +0000 2023	zelayapuspita24	@B3doel___ gerakan boikot produk pro zionis memang harus berjalan berkelanjutan biar efeknya maksimal. stop pakai danone aqua, mcd, starbucks, dan sejenisnya

### 4.2. Preprocessing Text

Pada tahapan ini terdiri dari 5 proses yaitu *case folding*, *data cleaning*, *normalization*, *stopwords removal*, dan *stemming*. Berikut merupakan contoh hasil dari *preprocessing text* pada tabel 3.

Tabel 3. Contoh hasil *preprocessing text*

Praproses	Teks
Data	@kompascom Banyak yg tidak suka dgn Israel, brangkat bro ke jalur Gaza angkat senjata biar terpuaskan emosinya Kepada Israel, atau boikot semua produk Yahudi, supaya kalian tdk kelihatan lagi di medsos
Case Folding & Data Cleaning	banyak yg tidak suka dgn israel brangkat bro ke jalur gaza angkat senjata biar terpuaskan emosinya kepada israel atau boikot semua produk yahudi supaya kalian tdk kelihatan lagi di medsos
Normalization	banyak yang tidak suka dengan israel berangkat bro ke jalur gaza angkat senjata biar terpuaskan emosinya kepada israel atau boikot semua produk yahudi supaya kalian tidak kelihatan lagi di media sosial
Stopwords Removal	tidak suka israel berangkat jalur gaza angkat senjata biar terpuaskan emosinya israel boikot produk yahudi tidak kelihatan media sosial
Stemming	tidak suka israel berangkat jalur gaza angkat senjata biar puas emosi israel boikot produk yahudi tidak lihat media sosial

### 4.3. Penyaringan Data

Penyaringan data dilakukan untuk menghapus data teks bahasa asing dan data duplikat. Dari hasil pengumpulan data, ditemukan bahasa asing seperti bahasa melayu, bahasa inggris, bahasa sunda, dan bahasa jawa.

Data teks dengan bahasa asing sebanyak 1.907. Kemudian, hasil pengecekan *tweet* yang duplikat ditemukan sebanyak 1.695 *tweet*. Jadi, data yang siap digunakan sebanyak 7.881 data. Dari data tersebut, 95% akan digunakan untuk skenario klasifikasi dan 5% dari dataset akan digunakan untuk validasi sistem.

#### 4.4. Pelabelan Data

Pada tahap ini dilakukan pelabelan sentimen positif dan negatif menggunakan Vader. Penentuan label ditentukan dari nilai *compound score*. Nilai *compound score* > 0.05, labelnya positif dan nilai *compound score* < -0.05, labelnya negatif. Dari hasil pelabelan, diperoleh data dengan sentimen positif sebanyak 5.407 dan sentimen negatif sebanyak 2.474. Berikut merupakan contoh data yang telah dilabeli pada tabel 4.

Tabel 4. Hasil pelabelan sentimen

Teks	Compound Score	Sentimen
demo boikot produk israel	1,23541666 666667	Positif
kalau boikot produk israel indonesia negara belakang selama pakai guna buat yahudi israel lampu listrik internet android facebook instagram google mobil sepeda motor bos orang yahudi	-0.0516	Negatif

#### 4.5. Pembagian Data

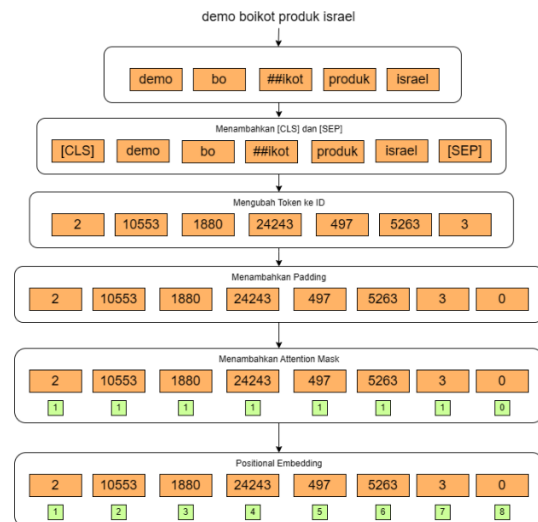
Dataset untuk klasifikasi dibagi menjadi 2 bagian yaitu *data train* sebesar 80% dan *data test* sebanyak 20%. Dataset yang telah diberikan label mengalami ketidakseimbangan kelas, maka perlu diseimbangkan terlebih dahulu menggunakan *random oversampling* yang mana kelas negatif akan disamakan jumlahnya dengan kelas positif. Sehingga jumlah dataset menjadi 10.256. Berikut hasil pembagian data latih dan data uji yang telah diseimbangkan ditunjukkan pada tabel 5.

Tabel 5. Pembagian data

Data Latih	Data Uji
8.204	2.052

#### 4.6. Klasifikasi

Sebelum dilakukan pelatihan, dilakukan *word embedding*. Hasil tahapan *word embedding* diilustrasikan pada gambar 4.



Gambar 4. Contoh ilustrasi *word embedding*

Tahap klasifikasi dilakukan dengan menggunakan IndoBERT ('indobenchmark/indobert-base-p1). Tahap klasifikasi akan dilakukan dalam 2 skenario yaitu sebagai berikut.

Tabel 6. Skenario klasifikasi

Skenario	Batch Size	Learning Rate	Dropout Rate
S1	16	2e-5	0.1
S2	32	2e-5	0.1

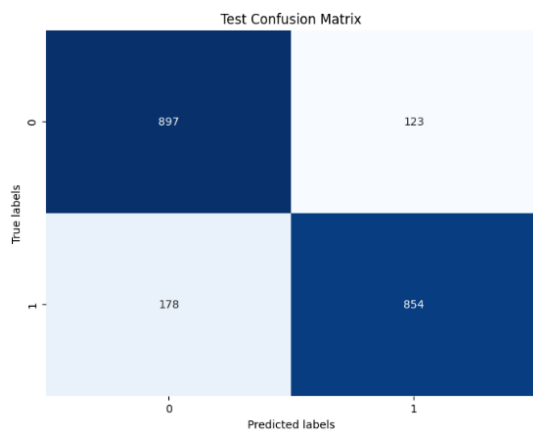
#### 4.7. Evaluasi

Berdasarkan hasil tahapan dari klasifikasi, dilakukan evaluasi dalam setiap skenario. Berikut merupakan hasil evaluasi dari setiap skenario yang ditunjukkan pada tabel 7.

Tabel 7. Hasil Evaluasi

Skena rio	Accur acy	Precis sion	Recall	F1- score
S1	0.853	0.854	0.854	0.853
S2	0.844	0.846	0.844	0.844

Dilihat dari tabel 6, nilai akurasi tertinggi didapatkan pada skenario 1 dengan nilai akurasi 85%, *precision* 85%, *recall* 85%, serta *f1-score* 85%. Berikut merupakan *confusion matrix* dari skenario 1.



Gambar 5. *Confusion matrix* skenario 1

Gambar 5 menunjukkan hasil evaluasi menggunakan *confusion matrix* pada skenario 1 yang mana model sudah cukup baik mengklasifikasikan setiap sentimen. Dari 2.052 data uji, model mampu memprediksi dengan benar sebanyak 1.751 data sentimen.

#### 4.8. Visualisasi dan Validasi

Visualisasi dilakukan dengan membuat sebuah *website* dengan *framework streamlit*. Visualisasi *website* ditunjukkan pada gambar 6.



Gambar 6. Visualisasi halaman *upload file csv*

Validasi sistem dilakukan dengan menggunakan data yang disisihkan sebelumnya dan menghasilkan akurasi 78%.

### 5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dikerjakan, dapat ditarik kesimpulan yaitu:

- Sentimen positif terhadap aksi boikot mengarah pada dukungan melakukan boikot dan sentimen negatif mengarah pada ketidaksetujuan boikot serta kerugian dari adanya aksi tersebut. Sentimen yang lebih banyak yaitu sentimen positif sehingga publik banyak yang setuju dengan adanya aksi boikot produk israel.

- Dari hasil percobaan, maka skenario 1 dengan *learning rate*  $2e-5$ , *batch size* 16, *dropout rate* 0.1 memiliki akurasi yang paling tinggi yaitu 85%. Nilai presisi 85%, nilai *recall* 85%, dan nilai *F1-score* 85%. Hasil validasi sistem mendapatkan akurasi 78%.
- Visualisasi *website* dengan *framework streamlit* yang menunjukkan grafik dan tren sentimen. Halaman untuk validasi dibuat dalam bentuk *input text* dan *upload file csv*.

### UCAPAN TERIMA KASIH

Penulis memberikan ucapan terima kasih kepada pihak yang telah yang mendukung penyelesaian dari penelitian ini.

### DAFTAR PUSTAKA

- [1] M. Alberto, I. Simarmata, and N. Yuliana, "Triwikrama: Jurnal Multidisiplin Ilmu Sosial Analisis Framing Konflik Israel-Palestina Dalam Sindonews.Com Dan Cnn Indonesia," vol. 2, pp. 2023–2054, 2023.
- [2] H. Sabiah Vitry, K. Ummatin, M. Hasni Azzahra, A. Putri Amanda, and D. Permata Suci, "Triwikrama: Jurnal Multidisiplin Ilmu Sosial Konflik Israel Dan Palestina 'Analisis Manajemen Konflik Yang Mempengaruhi Mental Health Anak Anak Palestina,'" vol. 2, no. 2, pp. 2023–2024, 2023.
- [3] M. Alfa, R. Mokobombang, F. Ayu, L. Niu, and J. Hasan, "Perilaku Boikot Dalam Perspektif Islam Serta Implementasinya Di Era Kontemporer," 2023. [Online]. Available: <http://ejournal.iain-manado.ac.id/index.php/maqrizii>
- [4] M. Risqi, F. Septiazi, and N. Yuliana, "Triwikrama: Jurnal Multidisiplin Ilmu Sosial Analisis Pengaruh Media Sosial Terhadap Gerakan Boikot Produk Israel Di Indonesia," vol. 2, no. 4, pp. 2023–2054, 2023.
- [5] H. Nanang, "Komparasi Algoritma Random Forest, Naïve Bayes, dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN) Nanang Husin," 2023.
- [6] D. della Porta and M. Diani, *Social Movements: An Introduction*. 2009.
- [7] B. Pang and L. Lee, *Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval*. 2008.
- [8] M. F. Y. Herjanto and C. Carudin, "Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma



- Random Forest Classifier,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4192.
- [9] E. Rosenberg, C. Tarazona, and R. Vinuesa, “Sentiment analysis on Twitter data towards climate action”, doi: 10.21203/rs.3.rs-2434092/v1.
- [10] S. Diantika, “Penerapan Teknik Random Oversampling Untuk Mengatasi Imbalance Class Dalam Klasifikasi Website Phishing Menggunakan Algoritma Lightgbm,” 2023.
- [11] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019. [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [12] I. Werdiningsih, D. C. R. Novitasari, and D. Z. Haq, *Pengelolaasn Data Mining dengan Pemrograman Matlab*. Surabaya: Airlangga University Press, 2022.