

# Penerapan Algoritma Cerdas Bidirectional Encoder Representations From Transformers Dalam Menganalisis Opini Publik Terhadap Produk Yang Mengalami Boikot

Asep Surahman Sulaeman<sup>1</sup>, Alun Sujjada<sup>2</sup>, Ivana Lucia Kharisma<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Universitas Nusa Putra

Jl. Raya Cibolang No.21 Cisaat, Sukabumi, Jawa Barat

E-mail: [asep.surahman\\_ti20@nusaputra.ac.id](mailto:asep.surahman_ti20@nusaputra.ac.id)<sup>1</sup>, [alun.sujjada@nusaputra.ac.id](mailto:alun.sujjada@nusaputra.ac.id)<sup>2</sup>,  
[ivana.lucia@nusaputra.ac.id](mailto:ivana.lucia@nusaputra.ac.id)<sup>3</sup>

**Abstract** – Instagram, as a social media platform, has become instrumental in voicing opinions and participating in public discussions on various social, economic, and political issues. One prominent issue discussed in this study is product boycotts, which can significantly impact brand image and sales. Products from well-known brands such as McDonald's, KFC, Starbucks, Burger King, and Pizza Hut have become primary targets in these boycott actions. This research aims to analyze how international public perception responds and understands the Israel-Palestine conflict, particularly regarding their support for these boycott actions. The study utilizes a dataset of 1,750 Instagram comments related to these products, classified into positive and negative labels through automatic transformer-based and manual labeling. Sentiment analysis results indicate that McDonald's has a positive sentiment of 41.43% and a negative sentiment of 58.57%, KFC with 85.14% positive and 14.86% negative, Starbucks with 97.71% positive and 2.29% negative, Burger King with 50% positive and negative, and Pizza Hut with 80.57% positive and 19.43% negative. Using pre-trained Bidirectional Encoder Representations from Transformers (BERT) model (Bert-Base-Uncased), accuracy results are obtained for McDonald's at 84.14%, KFC at 95%, Starbucks at 94.16%, Burger King at 91.42%, and Pizza Hut at 93.80%.

**Keywords** - Product Boycott, Pre-trained Bidirectional Encoder Representations From Transformers (BERT), Bert-base-uncased, Sentiment Analysis

**Intisari** – Media sosial Instagram, telah menjadi platform untuk menyampaikan pendapat dan berpartisipasi dalam diskusi publik mengenai berbagai isu sosial, ekonomi, dan politik. Salah satu isu yang ramai saat penelitian ini dibuat adalah tentang boikot produk. Memboikot suatu produk dapat berdampak signifikan terhadap citra merek dan penjualan. Produk dengan merek terkenal seperti *McDonald's*, *KFC*, *Starbucks*, *Burger King*, dan *Pizza Hut* menjadi target utama dalam aksi boikot. Hasil penelitian bertujuan untuk menganalisis bagaimana persepsi publik luar negeri merespons serta memahami konflik antara Israel-palestina dan sejauh mana dukungan mereka terhadap aksi boikot. Penelitian ini menggunakan 1.750 dataset yang diambil dari komentar akun Instagram pada produk terkait. Data terbagi menjadi 2 label positif dan negatif berdasarkan pelabelan otomatis dari transformer dan secara manual. Hasil analisis sentimen menunjukkan bahwa *McDonald's* memiliki sentimen positif sebesar 41,43% dan negatif 58,57%, *KFC* dengan 85,14% positif dan 14,86% negatif, *Starbucks* dengan 97,71% positif dan 2,29% negatif, *Burger King* dengan 50% positif dan negatif, serta *Pizza Hut* dengan 80,57% positif dan 19,43% negatif. Dari hasil pemodelan dengan metode *pre-trained Bidirectional Encoder Representations From Transformers* (BERT) dari Bert-Base-Uncased, diperoleh hasil *accuracy* pada produk *McDonald's* sebesar 84,14%, produk *KFC* 95%, produk *Starbucks* 94,16% produk *Burger King* 91,42%, dan produk *Pizza Hut* 93,80%.

**Kata Kunci** – Boikot Produk, *Pre-trained Bidirectional Encoder Representations From Transformers* (BERT), *Ber-base-uncased*, Analisis Sentimen

## I. PENDAHULUAN

Media social Instagram telah menjadi salah satu platform bagi masyarakat untuk menyampaikan pendapat dan berpartisipasi dalam diskusi publik. Pada sosial media, pengguna dapat dengan bebas berbagi serta memperoleh informasi, perkembangan sosial media yang pesat dapat mempengaruhi sikap dan pola perilaku [1]. Instagram, sebagai salah satu media sosial yang sangat populer di dunia, di mana pengguna sangat aktif berinteraksi, meninggalkan komentar, dan berbagi pendapat tentang berbagai topik, termasuk aksi boikot produk. Instagram berbasis visual menjadi tempat menarik dalam melakukan kampanye boikot karena pengguna cenderung memberikan tanggapan pada postingan terkait seperti like, komentar, dan share cenderung lebih tinggi dibandingkan beberapa platform lain, sehingga komentar-komentar tersebut bisa memberikan data sentiment yang beragam dalam sebuah komentar aksi boikot.

Boikot adalah tindakan untuk tidak menggunakan, membeli, atau berurusan dengan seseorang, organisasi, atau negara [2]. Berdasarkan Kamus Besar Bahasa Indonesia, boikot diartikan sebagai tindakan untuk menolak bekerja sama dalam berbagai hal seperti urusan dagang, atau berpartisipasi. Pemboikotan dilakukan sebagai bentuk protes dari satu pihak terhadap pihak lain yang dianggap melakukan tindakan yang tidak pantas, yang biasanya diwujudkan dalam bentuk penolakan [3]. Boikot produk ini dilakukan karena dugaan bahwa sebagian pendapatan digunakan untuk membantu Israel, termasuk bantuan dana infrastruktur dan pengembangan senjata yang menguatkan posisi Israel di Palestina. Oleh karena itu, memboikot produk merupakan dukungan moral terhadap perjuangan rakyat Palestina sekaligus protes terhadap tindakan yang mendukung serangan Israel terhadap Palestina. Dengan harapan menurunnya permintaan atas barang dan jasa produk boikot di pasar lokal maupun internasional. Memboikot suatu produk dapat berdampak signifikan terhadap citra merek dan penjualan. Produk dengan merek-merek terkenal seperti *McDonald's*, *KFC*, *Starbucks*, *Burger King*, dan *Pizza Hut* membuatnya menjadi target utama dalam aksi boikot. Pemilihan produk tersebut didasarkan pada peran signifikan merek-merek tersebut dalam globalisasi ekonomi, di mana merek tersebut sangat signifikan terhadap perekonomian Israel, termasuk dana untuk pembangunan infrastruktur dan pengembangan senjata yang mengancam Negara Palestina.

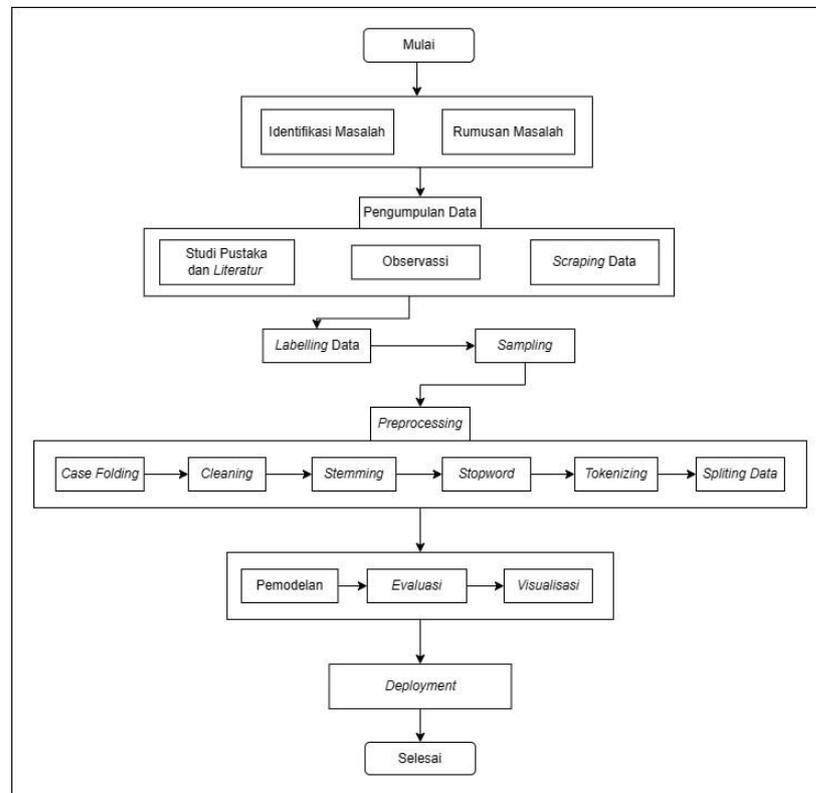
Analisis sentimen merupakan bagian dari *text mining* yang mempelajari komputasi opini orang-orang, emosi, dan sentimen dengan tujuan untuk mengklasifikasikan sentimen dalam suatu kalimat yang berisi opini. Analisis ini dapat mengidentifikasi apakah kalimat memiliki sentimen positif atau negatif berdasarkan polaritasnya [5]. Pada proses penelitian ini menggunakan Algoritma *Bidirectional Encoder Representations from Transformers* (BERT).

BERT adalah representasi encoder dari model *Transformers*, sebuah arsitektur NLP yang menggantikan jaringan berulang dengan mekanisme perhatian untuk menangkap hubungan kontekstual antara kata-kata yang berjauhan. BERT adalah model bidirectional yang membaca teks dari kedua arah (kiri ke kanan dan kanan ke kiri) secara bersamaan, yang memberikan pemahaman konteks yang lebih baik dibandingkan dengan model satu arah seperti RNN atau LSTM. BERT dapat memproses sebuah kata pada kalimat berdasarkan ada atau tidaknya kaitan antara kata tersebut dengan kalimat secara keseluruhan. Selain itu, BERT dilatih sebelumnya (*pre-trained*) pada kumpulan data yang sangat besar dan beragam, sehingga memiliki pengetahuan yang luas tentang bahasa. Dengan fine-tuning pada dataset seperti komentar Instagram tentang boikot produk, BERT dapat mencapai kinerja yang sangat baik dengan sedikit data pelatihan dibandingkan metode lain yang harus dilatih dari awal. BERT juga menggunakan teknik tokenisasi berbasis sub-kata (*sub-word tokenization*), yang membuatnya lebih efektif dalam menangani kata-kata yang tidak dikenal. Studi telah menunjukkan bahwa BERT secara konsisten mengungguli model-model lain dalam berbagai tugas NLP, termasuk klasifikasi teks, penarikan informasi, dan analisis sentimen.

## II. SIGNIFIKASI STUDI

### A. Tahapan Penelitian

Tahapan penelitian mencakup langkah-langkah yang dilakukan dari awal hingga akhir penelitian. Adapun langkah-langkah tersebut dapat dilihat pada gambar 1 berikut:



Gambar 1. Tahapan Penelitian

Langkah awal menentukan topik penelitian dengan identifikasi serta rumusan masalah. Proses pengumpulan data, data akan diberikan label positif dan negatif. Diketahui jumlah label positif dan negatif tidak seimbang, maka *sampling* perlu dilakukan, masuk proses preprocessing. *splitting* data, tahap pemodelan, di evaluasi, *visualisasi* serta di deployment kedalam website.

### B. Identifikasi Masaalah

Boikot dilakukan sebagai bentuk protes ketidakpuasan dari satu pihak terhadap pihak lain yang dianggap melakukan tindakan yang tidak semestinya, biasanya diwujudkan melalui tindakan penolakan [3]. Produk dari merek terkenal seperti *McDonald's*, *KFC*, *Starbucks*, *Burger King*, dan *Pizza Hut* menjadi fokus utama dalam aksi boikot. Mengklasifikasikan setiap komentar secara manual memerlukan waktu lama dan tidak efektif, sehingga diperlukan sistem untuk menganalisis komentar terkait boikot produk..

Rumusan masalah pada penelitian ini bagaimana implementasi metode BERT untuk analisis sentiment komentar Instagram pada akun *McDonald's*, *KFC*, *Starbucks*, *Burger King*, dan *Pizza Hut*.

### C. Metode Pengumpulan Data

Pada penelitian ini metode yang digunakan yaitu metode kuantitatif deskriptif yang melibatkan observasi serta pengumpulan data untuk menganalisis populasi atau sampel tertentu dengan mengumpulkan data menggunakan instrumen tertentu. Metode kuantitatif digunakan

untuk melakukan evaluasi numerik terhadap kemampuan BERT dalam mengklasifikasikan komentar, dengan menggunakan skala atau nilai yang dapat diukur.

### 1. *Studi Pustaka dan Literatur*

Peneliti melakukan studi pustaka dengan mempelajari berbagai buku, skripsi, jurnal, serta mengunjungi beberapa website dan tutorial terkait analisis sentimen, preprocessing, dan algoritma BERT.

### 2. *Observasi*

Peneliti melakukan observasi dengan metode pengamatan dan pengumpulan data komentar melalui platform media Instagram terkait dengan boikot produk.

### 3. *Scraping*

Pengumpulan data komentar dilakukan dengan alat *Scraping* di browser *Google Chrome* dengan *Data Miner*. Alat ini berfungsi untuk mengambil data dari website dan menyimpannya dalam format *Excel* atau *csv*. Data komentar yang diambil berjumlah 1.750, dengan masing-masing 350 data pada produk boikot di luar negeri.

### D. *Labelling Data*

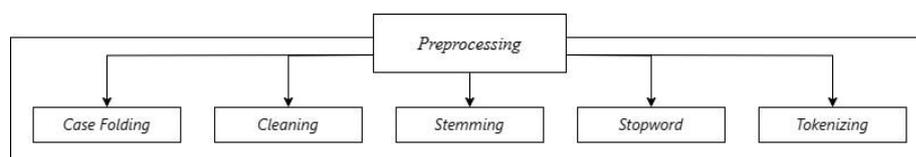
Pelabelan dilakukan secara otomatis menggunakan *Natural Language Processing (NLP)* yang ada pada *library transformers*. Dataset yang telah diberi label secara otomatis akan diperiksa kembali untuk memastikan apakah labelnya sudah benar atau perlu diperbaiki lagi secara manual. Pada penelitian ini komentar yang berabel positif merupakan komentar yang tidak ada hubungannya dalam boikot produk, menyukai, memuji baik dari segi rasa, harga maupun kepuasannya. Dan berlabel negatif berisikan tentang ketidaksetujuan yang menyatakan penolakan terhadap produk-produk tersebut.

### E. *Sampling*

Dataset yang sudah memiliki label akan dihitung, jika pada jumlah dataset menunjukkan ketidakseimbangan maka proses *sampling* perlu diperlukan agar klasifikasi dilakukan secara akurat dan tidak memihak terhadap sampel mayoritas. [14]. Dalam penelitian ini, menggunakan teknik *Random Over Sampling*, di mana data dari sampel minoritas akan diduplikasi sehingga jumlahnya setara dengan sampel mayoritas.

### F. *Preprocessing*

*Preprocessing* merupakan sebuah tahapan pembersihan data agar dapat digunakan untuk melakukan proses *training* data. Langkah ini mempermudah pengenalan data oleh komputer [14]. *Preprocessing* terbagi menjadi beberapa tahapan, seperti pada gambar 2 yang mencakup *Case Folding*, *Cleaning*, *Stemming*, *Stopword*, dan *Tokenizing*.



Gambar 2. *Preprocessing*

Beberapa proses diantaranya *Case Folding* mengubah semua huruf menjadi huruf kecil (*lowercase*), *Cleaning* membersihkan data dan noise, *Stemming* mengubah kata menjadi bentuk dasarnya *Stopword* menghapus kata-kata yang tidak memiliki makna, serta *Tokenizing* membagi teks menjadi token-token yang mewakili unit-unit terkecil.

### G. Splitting

*Splitting Data* melibatkan pembagian dataset menjadi dua bagian: data latih dan data uji. Data latih digunakan untuk melatih model, sementara data uji digunakan untuk menguji model. Pada tahap ini, pembagian data menggunakan proporsi 80:20, menunjukkan bahwa 80% dari keseluruhan data akan digunakan sebagai data *training-validation* (60%-20%), sementara 20% sisanya akan digunakan sebagai data *testing* [17]. Data *validation* dimanfaatkan selama proses pelatihan untuk mencegah *overfitting*.

### H. Pemodelan

Pada penelitian ini menggunakan metode BERT. BERT adalah model representasi bahasa yang menghasilkan model *pre-train representasi bidirectional* dari teks yang tidak berlabel dengan memperhitungkan kedua konteks dari setiap layer [19]. BERT dilatih (*pre-trained*) pada kumpulan data yang sangat besar dan beragam, sehingga memiliki pengetahuan yang luas tentang bahasa. Dengan *fine-tuning* pada dataset, BERT dapat mencapai kinerja yang sangat baik dengan sedikit data pelatihan dibandingkan metode lain yang harus dilatih dari awal. Pada penelitian ini model BERT yang digunakan adalah model "*bert-base-uncased*". *BERT-base-uncased* adalah model yang dilatih menggunakan data teks yang besar dan dirancang untuk memahami konteks kalimat karena kemampuannya yang unggul dalam memahami konteks tanpa memperhatikan huruf besar atau kecil dalam teks.

### I. Evaluasi

*Evaluasi* dilakukan untuk memperoleh pemahaman sejauh mana model atau metode berhasil mengevaluasi dalam memprediksi data yang belum pernah terlihat sebelumnya. Dalam penelitian ini, evaluasi akurasi menggunakan confusion matriks:

#### 1. Akurasi

Akurasi merupakan seberapa akurat atau sejauh mana model mampu memprediksi dengan benar terhadap keseluruhan data.

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

#### 2. Precision

*Precision* adalah perbandingan antara jumlah prediksi positif yang tepat dengan total prediksi positif yang dilakukan.

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

#### 3. Recall

*Recall* merupakan proporsi hasil prediksi positif yang tepat dibandingkan dengan total jumlah observasi positif yang sebenarnya dalam suatu kelas tertentu.

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

#### 4. F1 Score

*F1 Score* merupakan nilai rata-rata yang memberikan bobot pada *precision* dan *recall*.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{recall}}{\text{Precision} + \text{recall}} \quad (4)$$

### J. Visualisasi

Pada langkah ini, informasi data akan disajikan secara visual untuk mudah dipahami. Teknik *visualisasi* yang digunakan adalah *wordcloud*, yang dapat menampilkan kata-kata yang paling sering muncul dengan ukuran yang berbeda sesuai dengan frekuensinya.

### K. Deployment

Model yang sudah terlatih akan dijalankan sebagai aplikasi web menggunakan *framework Streamlit*. Aplikasi ini memiliki *User Interface* (UI) yang memungkinkan pengguna untuk melakukan input dan akan menerima hasil prediksi analisis yang dihasilkan oleh model.

### III. HASIL DAN PEMBAHASAN

#### A. Dataset

Dataset berupa sebuah komentar dari beberapa postingan akun Instagram *McDonalds*, *KFC*, *Starbucks*, *Burger King*, *Pizza Hut*, dengan Pengumpulan data (*Scraping*) menggunakan Data Miner. Total data yang berhasil dikumpulkan adalah sebagai berikut:

TABEL I  
HASIL PENGUMPULAN DATA

No	Akun Instagram	Jumlah Postingan yang diambil	Data yang diambil	Jumlah Data keseluruhan
1	<i>McDonald's</i>	1 Postingan pada tanggal 2 Maret 2024	350 Data	
2	<i>KFC</i>	2 Postingan pada tanggal 2 dan 5 Maret 2024	350 Data	
3	<i>Starbucks</i>	2 Postingan pada tanggal 19 dan 26 Maret 2024	350 Data	1.750 Dataset
4	<i>Burger King</i>	2 Postingan pada tanggal 8 dan 9 Maret 2024	350 Data	
5	<i>Pizza Hut</i>	3 Postingan pada tanggal 4, 6 dan 11 Maret 2024	350 Data	

Diketahui pengambilan dataset menggunakan Teknik *Scraping* Data Miner dilakukan pengambilan pada beberapa postingan akun produk boikot pada bulan Maret 2024 masing-masing produk sebanyak 350 data.

#### B. Labelling

Dataset akan diberikan label agar model dapat mengenali dan mengidentifikasi pola atau karakteristik. Label dilakukan secara otomatis menggunakan *library Transformers*. Dataset dibagi menjadi dua kategori yaitu: berlabel positif dan negatif. Berikut merupakan jumlah label positif dan label negatif secara otomatis:

TABEL II  
HASIL LABELLING OTOMATIS

No	Akun Instagram	Label	
		Positif	Negatif
1	<i>McDonald's</i>	78 Data	272 Data
2	<i>KFC</i>	102 Data	248 Data
3	<i>Starbucks</i>	119 Data	231 Data
4	<i>Burger King</i>	99 Data	251 Data
5	<i>Pizza Hut</i>	100 Data	250 Data
Jumlah Keseluruhan Data		498 Data	1.252 Data
		1.750 Data	

Beberapa data ditemukan tidak sesuai dalam pelabelan, oleh karena itu perlu di cek kembali secara manual. Diketahui setelah dilakukan perbaikan pelabelan secara manual mengakibatkan perubahan pada jumlah data positif dan negatif. Seperti yang ditampilkan pada tabel berikut:

TABEL III  
HASIL LABELLING MANUAL

No	Akun Instagram	Label	
		Positif	Negatif
1	<i>McDonald's</i>	145 Data	205 Data
2	<i>KFC</i>	52 Data	298 Data
3	<i>Starbucks</i>	342 Data	8 Data

No	Akun Instagram	Label	
		Positif	Negatif
4	<i>Burger King</i>	175 Data	175 Data
5	<i>Pizza Hut</i>	68 Data	282 Data
Jumlah Keseluruhan Data		746 Data	1.004 Data
		1.750 Data	

Diketahui setelah dilakukan pelabelan secara manual pada dataset, dataset memiliki nilai label positif dan negatif ketidakseimbangan data. Untuk mendapatkan akurasi yang lebih tinggi pada model, maka perlu dilakukan proses *Random Oversampling*.

C. *Sampling*

Sebelum dilakukan *Random Oversampling* dataset masing-masing produk hanya memiliki 350 data, agar dapat menyeimbangkan antara jumlah label positif dan jumlah label negatif diperlukan proses *Random Oversampling*. Diketahui jumlah label positif dan negatif pada dataset tidak seimbang. Oleh karena itu, dilakukan *random oversampling*. Setelah proses *random oversampling*, dataset memiliki jumlah perubahan data sebagai berikut:

TABEL IV  
SETELAH *RANDOM OVERSAMPLING*

No	Akun Instagram	Data Setelah <i>Random Oversampling</i>		Jumlah data <i>Random Oversampling</i>
		Positif	Negatif	
1	<i>McDonald's</i>	205 Data	205 Data	410 Data
2	<i>KFC</i>	298 Data	298 Data	596 Data
3	<i>Starbucks</i>	342 Data	342 Data	684 Data
4	<i>Burger King</i>	175 Data	175 Data	350 Data
5	<i>Pizza Hut</i>	282 Data	282 Data	564 Data
Jumlah Keseluruhan Data		1.302 Data	1.302 Data	2.604 Data

Diketahui pada dataset setelah dilakukan *random oversampling* memiliki jumlah total data sebanyak 2.604, dengan jumlah nilai label positif dan negatif yang sama yaitu sebanyak 1.302 data.

D. *Preprocessing*

*Preprocessing* ini dibagi menjadi beberapa tahapan, diantaranya *Case Folding*, *Cleaning*, *Stemming*, *Stopword*, dan *Tokenizing*. Berikut merupakan tahapan-tahapan dari *Preprocessing*:

1. *Case Folding*

*Case folding* mengubah semua huruf menjadi huruf kecil (*lowercase*).

2. *Cleaning*

*Cleaning* dilakukan untuk menghilangkan noise pada dataset. Proses ini menghapus tanda baca, username, angka, hashtag, link, dan kata-kata yang tidak relevan.

3. *Stemming*

*Stemming* adalah proses mengubah kata ke bentuk dasarnya. Penelitian ini menggunakan NLTK (*Natural Language Toolkit*) untuk *stemming*, memudahkan dalam memproses kata-kata berbahasa Inggris dalam analisis teks.

4. *Stopword*

*Stopword* merupakan kata-kata umum yang tidak memiliki arti penting. Kata-kata yang dihilangkan yaitu: - penghubung seperti: and, or, in, to dan lain-lain.

5. *Tokenizing*

Tokenisasi pemisahan teks menjadi bagian yang disebut token. Bertujuan untuk memisahkan setiap kata dalam kalimat sehingga membentuk token-token terpisah [1]. Pada

proses tokenisasi ini, akan ditambahkan token khusus yang terdapat dalam kosa kata model. [CLS] ditambahkan di bagian awal untuk memberi tahu model. [SEP] untuk menandai akhir kalimat dan [PAD] sebagai *padding* menyesuaikan panjang data.

Hasil *Preprocessing* yang dilakukan melalui tahapan *Case Folding*, *Cleaning*, *Stemming*, *Stopword*, dan *Tokenizing*. Pada data komentar Instagram produk boikot ditunjukkan pada tabel berikut:

TABEL V  
PREPROCESSING

Preprocessing	Sebelum	Sesudah
<i>Case Folding</i>	<i>That's it I'm going to kfc now</i>	<i>that's it i'm going to kfc now</i>
<i>Cleaning</i>	<i>that's it i'm going to kfc now</i>	<i>thats it im going to kfc now</i>
<i>Stemming</i>	<i>thats it im going to kfc now</i>	<i>that it im go to kfc now</i>
<i>Stopword</i>	<i>that it im go to kfc now</i>	<i>im go kfc</i>
<i>Tokenizing</i>	<i>im go kfc</i>	<i>['[CLS]', 'im', 'go', 'k', '##fc', '[SEP]', '[PAD]' '[PAD]']</i>
<i>Encode</i>	<i>['[CLS]', 'im', 'go', 'k', '##fc', '[SEP]', '[PAD]' '[PAD]']</i>	<i>[101, 10189, 10271, 10211, 11783, 10114, 179, 10575, 10350, 11858, 102, 0]</i>
<i>Attention Mask</i>	<i>[101, 10047, 2175, 1047, 11329, 102, 0]</i>	<i>[1, 1, 1, 1, 1, 1, 0]</i>

Diketahui pada tabel diatas dilakukan proses *preprocessing* dari awal sampai akhir agar dataset dapat digunakan dalam melakukan proses *training* data serta mempermudah pengenalan data oleh komputer.

E. *Splitting Data*

Pembagian data menggunakan proporsi 80:20, menunjukkan 80% digunakan sebagai data *training-validation* (60%-20%), sementara 20% digunakan sebagai data *testing* [17]. Berikut merupakan pembagian data yang dilakukan:

TABEL VI  
PEMBAGIAN DATA

No	Akun Instagram	Nama Data			Jumlah Data
		Data Latih	Data Test	Data Validasi	
1	<i>McDonald's</i>	246	82	82	410 Data
2	<i>KFC</i>	357	120	119	596 Data
3	<i>Starbucks</i>	410	137	137	684 Data
4	<i>Burger King</i>	210	70	70	350 Data
5	<i>Pizza Hut</i>	338	113	113	564 Data
Jumlah Data					2.604

Diketahui pada data *training* memiliki data untuk dilatih yang lebih banyak, dikarenakan untuk memastikan model mendapatkan cukup informasi selama proses pelatihan sehingga dapat belajar dan membuat prediksi yang lebih akurat.

F. *Pemodelan*

Pada Tahapan ini data yang sudah siap akan dimasukan kedalam *model Pretrained BERT* dari "*bert-base-uncased*" dengan parameter sebagai berikut :

- a. Epochs : 5
- b. BATH\_SIZE : 32
- c. LEARNING\_RATE : 5e-5

Setelah itu, dilakukan konfigurasi *optimizer* menggunakan *optimizer ADAM*. Gambar dibawah menampilkan hasil dari pelatihan data produk boikot menggunakan *model Pretrained BERT*:

TABEL VII  
HASIL PELATIHAN MODEL

No	Akun Instagram	Model	
		Train Accuracy	Val Accuracy
1	McDonald's	95%	91%
2	KFC	96%	95%
3	Starbucks	94%	92%
4	Burger King	95%	87%
5	Pizza Hut	92%	91%

Dari hasil pelatihan menggunakan *model Pretrained BERT* pada masing-masing produk boikot, hasil nilai *Train accuracy* dan hasil nilai *Val accuracy* mendapatkan nilai yang cukup memuaskan.

G. Evaluasi

Setelah pelatihan model selesai, model tersebut akan dievaluasi untuk mengukur sejauh mana performanya terhadap data yang tidak terlihat sebelumnya. Pengujian ini dilakukan menggunakan *confusion matrix*, yang memungkinkan perhitungan beberapa metrik evaluasi performa model seperti yang diperlihatkan di bawah ini:

TABEL VIII  
PERFORMA MODEL EVALUASI

No	Akun Instagram	Positive			Negative		
		Precssion	Recall	F1_Score	Precssion	Recall	F1-Score
1	McDonald's	0.78	0.95	0.86	0.94	0.72	0.82
2	KFC	0.98	0.91	0.94	0.93	0.98	0.96
3	Starbucks	1.00	0.90	0.95	0.88	1.00	0.93
4	Burger King	0.90	0.95	0.92	0.93	0.88	0.90
5	Pizza Hut	0.90	0.98	0.94	0.98	0.89	0.94

Diketahui pada gambar diatas merupakan hasil pengujian performa model evaluasi menggunakan *confussion matrix* pada perhitungan matrik evaluasi pada masing-masing produk boikot.

Berdasarkan hasil yang sudah dilakukan, diperoleh perbandingan nilai akurasi yang menunjukkan performa model dalam mengklasifikasikan data. Hasil perbandingan nilai akurasi dapat dilihat pada tabel berikut:

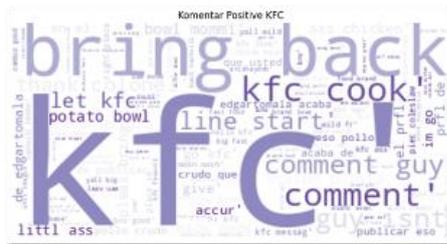
TABEL IX  
NILAI AKURASI PERBANDINGAN PRODUK BOIKOT

No	Akun Instagram	Nilai Akurasi
1	McDonald's	84,14%
2	KFC	95%
3	Starbucks	94,16%
4	Burger King	91,42%
5	Pizza Hut	93,80%

Dari hasil perhitungan nilai akurasi algoritma BERT pada produk *McDonald's* sebesar 84,14%, pada produk *KFC* sebesar 95%, produk ketiga *Starbucks* dengan nilai akurasi sebesar 94,16%, pada produk *Burger King* sebesar 91,42%, dan pada produk *Pizza Hut* sebesar 93,80%.

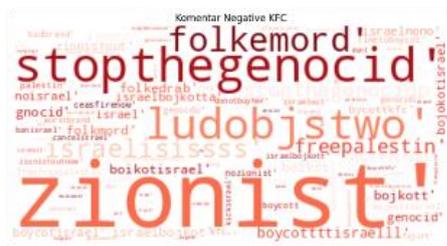
**H. Visualisasi**

*Visualisasi wordcloud* memberikan gambaran tentang kata-kata yang paling sering muncul dalam sebuah komentar. Berikut ini adalah hasil dari visualisasi wordcloud yang telah dilakukan



Gambar 3. Positive

Diketahui pada gambar diatas menunjukkan kata yang paling muncul adalah kata "kfc", dan "bring back". Menunjukkan bahwa pada sentiment positif sebagian besar diskusi publik membahas tentang kata "kfc".



Gambar 4. Negative

Pada gambar diatas diketahui kata yang paling muncul adalah kata "zionist", "stopthegenocid", dan "folkemord". Menunjukkan sebagian besar diskusi publik pada sentimen negatif berfokus pada isu yang terkait dengan *Zionisme*.

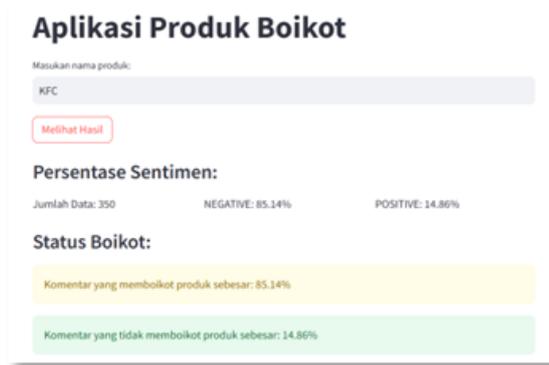
**I. Deployment**

Hasil modeling yang sudah dilakukan sebelumnya menggunakan BERT diintegrasikan ke dalam sebuah sistem menggunakan *framework streamlit*. Tujuannya agar model memiliki *user interface* yang memungkinkan pengguna untuk melakukan input. Seperti pada Gambar berikut:



Gambar 5. Tampilan Home UI

Tampilan menu *Home* yang merupakan halaman utama serta memprediksi status produk boikot. Pada halaman *McDonald's*, *KFC*, *Starbucks*, *Burger King* dan *Pizza Hut* merupakan penelitian yang sudah dilakukan sebelumnya. Aplikasi produk boikot memudahkan pengguna dalam melakukan analisis data sentimen. Ketika pengguna menginputkan data produk kedalam aplikasi menghasilkan output status produk boikot, komentar yang memboikot produk dan komentar yang tidak memboikot produk. Berikut merupakan hasil yang didapat:



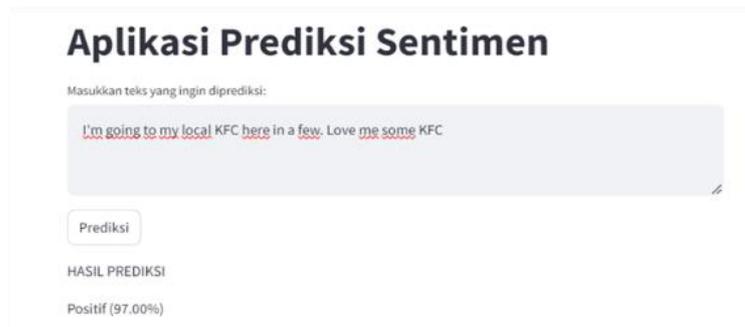
Gambar 6. Status Produk *KFC*

Diketahui pada produk *KFC* memiliki jumlah data sebanyak 350 data. Komentar yang memboikot produk *KFC* sebesar 85,14% dan komentar yang tidak memboikot produk *KFC* sebesar 14,86%. Berikut merupakan status produk boikot dari masing-masing produk:

TABEL X  
STATUS PRODUK BOIKOT

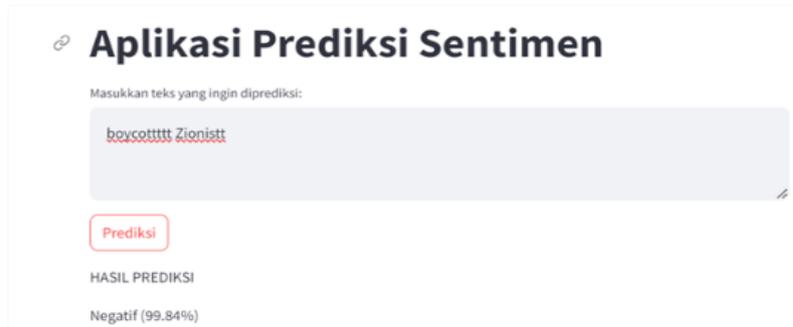
No	Akun Instagram	Komentar Yang Memboikot Produk	Komentar Yang Tidak Memboikot Produk
1	<i>McDonald's</i>	58,57%	41,47%
2	<i>KFC</i>	85,14%	14,86%
3	<i>Starbucks</i>	2,29%	97,71%
4	<i>Burger King</i>	50%	50%
5	<i>Pizza Hut</i>	80,57%	19,43%

Pada tabel diatas diketahui bahwa pada masing-masing produk memiliki nilai status boikot yang berbeda, nilai komentar yang memboikot produk lebih tinggi pada produk *KFC*. Sementara nilai komentar yang memboikot produk paling rendah adalah *Starbucks*. Tampilan *User Interface* yang dibuat pada aplikasi berbasis website ini merupakan hasil pemodelan yang dapat melakukan prediksi terhadap teks komentar yang dimasukkan. Berikut merupakan tampilan *user interface* yang dihasilkan dari pemodelan:



Gambar 7. Prediksi Komentar Positif

Pada gambar tersebut diketahui ketika pengguna memasukan teks komentar “*I’m going to my local KFC here in a few. Love me some KFC*” prediksi yang dilakukan oleh model bernilai positif dengan nilai akurasi sebesar 97.00%.



Gambar 8. Prediksi Komentar Negatif

Diketahui pada gambar diatas ketika pengguna memasukan teks komentar “*boycottttt Zionistt*” prediksi yang dilakukan oleh model cukup memuaskan yaitu dengan bernilai negative dengan nilai akurasi sebesar 99.84%.

#### IV. KESIMPULAN

Pengambilan data menggunakan metode *Scraping* Data Miner menghasilkan 10 file dataset Excel dan berjumlah 1.750 dataset komentar, Pada produk *McDonald’s* komentar yang memboikot produk sebesar 58,57% dan komentar yang tidak memboikot sebesar 41,43% menunjukkan bahwa produk tersebut memiliki sentiment relatif lebih tinggi dalam aksi boikot. Pada produk *KFC* komentar yang memboikot sebesar 85,14% dan komenar yang tidak memboikot sebesar 14,86% tingkat boikot yang tinggi menunjukkan bahwa produk tersebut memiliki dukungan yang kuat dalam aksi boikot. Pada produk *Starbucks* komentar yang memboikot sebesar 2,29% dan komentar yang tidak memboikot sebesar 97,71% menunjukkan bahwa produk tersebut hanya memiliki sedikit dukungan dalam menyuarakan aksi boikot. Pada produk *Burger King* komentar yang memboikot sebesar 50% dan komentar yang tidak memboikot produk sebesar 50% menunjukkan bahwa produk tersebut memiliki kestabilan dalam aksi boikot. Pada produk *Pizza Hut* komentar yang memboikot sebesar 80,57% dan yang tidak tidak memboikot sebesar 19,43% tingkat boikot yang tinggi menunjukkan bahwa produk tersebut memiliki dukungan yang kuat dalam aksi boikot. Tingkat akurasi persentase keberhasilan BERT dalam melakukan klasifikasi komentar pada produk *McDonald’s* sebesar 84,14%, pada produk *KFC* sebesar 95% pada produk *Starbucks* sebesar 94,16%, pada produk *Burger King* sebesar 91,42% dan pada produk *Pizza Hut* sebesar 93,80%.

#### REFERENSI

- [1] A. Nayla, C. Setianingsih, and B. Dirgantoro, “Deteksi Hate Speech Pada Twitter Menggunakan Algoritma BERT,” *e-Proceeding Eng.*, vol. 10, no. 1, pp. 256–262, 2023.
- [2] Oktavia, M. R. Noval, R. Hanipah, and M. F. Handayani, “Pengaruh Dampak Boikot Produk Amerika Terhadap Perokonomian Indonesia,” *J. Mutiara Ilmu Akunt.*, vol. 2, no. 1, pp. 318–323, 2024, [Online]. Available: <https://doi.org/10.55606/jumia.v2i1.2377>
- [3] N. Elia, “Strategi Public Relations : Membangun Citra Perusahaan Terhadap Produk Boikot ARTIKEL ILMIAH Strategi Public Relations : Membangun Citra Perusahaan Terhadap Produk Boikot Pro-Israel Dosen Pengampu Alifa Nur Fitri , M . I . Kom Disusun Oleh,” no. December, pp. 0–17, 2023.
- [4] A. Tiara Susilawati, A. H. Tiara Susilawati Universitas Muhammadiyah Kalimantan Timur Nur Anjeni Lestari Universitas Muhammadiyah Kalimantan Timur Puput Alpria Nina

- Universitas Muhammadiyah Kalimantan Timur Jl Ir Juanda No, K. Samarinda Ulu, K. Samarinda, and K. Timur, "Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naïve Bayes," *J. Ilm. Mhs.*, vol. 2, no. 1, pp. 26–35, 2024, [Online]. Available: <https://doi.org/10.59603/niantanasikka.v2i1.240>
- [5] E. Y. Hidayat, R. W. Hardiansyah, and A. Affandy, "Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 108–118, 2021, doi: 10.25077/teknosi.v7i2.2021.108-118.
- [6] Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *J. Buana Inform.*, vol. 14, no. 02, pp. 107–116, 2023, doi: 10.24002/jbi.v14i02.7244.
- [7] "No Title." [Online]. Available: <https://www.dicoding.com/blog/apa-itu-data-mining/#:~:text=Data mining adalah proses pengumpulan dan pengolahan data,perhitungan statistika%2C matematika%2C ataupun teknologi Artificial Intelligence %28AI%29>.
- [8] K. M. Rahmi, "Jurnal 5 TEXT MINING ANALYSIS DAN SENTIMENT ANALYSIS DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER(Studi Kasus: Data Tanggapan Mengenai Tokopedia Melalui Media Sosial Twitter)," pp. 1–84, 2021.
- [9] A. Firdaus, W. I. Firdaus, P. Studi, T. Informatika, M. Digital, and P. N. Sriwijaya, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah," vol. 13, no. 1, pp. 66–78, 2021.
- [10] H. Ramanizar, A. Fajri, R. Binsar Sinaga, H. Mubarak, A. D. Pangestu, and D. S. Prasvita, "Analisis Sentimen Pengguna Twitter terhadap Konflik antara Palestina dan Israel Menggunakan Metode Naïve Bayesian Classification dan Support Vector Machine," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 166–175, 2021.
- [11] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4704–4713, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] A. Munandar and R. A. Firdaus, "Analisis Sentimen Netizen Indonesia Mengenai Boikot Produk," vol. 3, no. 1, pp. 23–40, 2023.
- [13] H. P. Utami, I. I. Prawisti, and S. M. Par, "Dampak Boikot Produk Makanan Amerika Terhadap Perkembangan Produk Makanan Lokal Indonesia," no. 229.
- [14] A. Dana, "Implementasi Metode *Bidirectional Encoder Representations from Transformers* (BERT) untuk Analisis Sentimen Komentar Pengguna," vol. 3, no. 1, 2023.
- [15] M. Yunus, "Basic Text *Preprocessing* menggunakan NLTKo Title," [yunusmuhammad007.medium.com/](https://yunusmuhammad007.medium.com/). Accessed: Mar. 23, 2024. [Online]. Available: <https://yunusmuhammad007.medium.com/basic-text-preprocessing-menggunakan-nltk-86ba3e65a1dc>
- [16] I. Kurniasari, H. Al Fatta, and Kusri, "Analisis Sentimen Opini Publik pada Instagram mengenai Covid-19 dengan SVM," *JTECS J. Sist. Telekomun. Elektron. Sist. Kontrol Power Sist. Komput.*, vol. 1, no. 1, pp. 67–74, 2021.
- [17] R. Darmawan and S. Amini, "Perbandingan Hasil Sentimen Analysis Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor pada Twitter Comparison of Sentiment Analysis Results Using Naïve Bayes and K-Nearest Neighbor Algorithm on Twitter," *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 495–501, 2022, [Online]. Available: <https://senafiti.budiluhur.ac.id/index.php/>
- [18] C. A. Putri, "Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan *Bidirectional Encoder Representations from Transformers*," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 2, pp. 181–193, 2020, doi: 10.35957/jatisi.v6i2.206.

- [19] N. Husin, “Komparasi Algoritma *Random Forest*, Naïve Bayes, dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN),” *J. Esensi Infokom J. Esensi Sist. Inf. dan Sist. Komput.*, vol. 7, no. 1, pp. 75–84, 2023, doi: 10.55886/infokom.v7i1.608.
- [20] L. S. Pradana, “Analisis sentimen masyarakat media sosial twitter terhadap kinerja pejabat gubernur Dki Jakarta menggunakan model indobert,” *Repository.Uinjkt.Ac.Id*, 2024, [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/77071%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/77071/1/LAMBANG SURYA PRADANA-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/77071%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/77071/1/LAMBANG%20SURYA%20PRADANA-FST.pdf)