


# Analisis Sentimen pada Ulasan Produk di Platform Perdagangan Digital (*E-Commerce*) Indonesia

Nadira Eka Rahmaharva<sup>a,1,\*</sup>

<sup>a</sup> Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia

<sup>1</sup> nadirae25@gmail.com\*

\* corresponding author

ARTICLE INFO		ABSTRACT
	NIM 2206051525	<p>Perkembangan <i>e-commerce</i> di Indonesia yang pesat menghasilkan <i>volume</i> ulasan produk digital dalam jumlah besar, sehingga diperlukan metode otomatis untuk analisis sentimen. Penelitian ini bertujuan mengembangkan model klasifikasi sentimen berbasis BERT (<i>Bidirectional Encoder Representations from Transformers</i>) untuk menganalisis ulasan produk pada platform <i>e-commerce</i> Indonesia. Dataset terdiri dari 11585 ulasan berbahasa Inggris (hasil terjemahan) dengan label sentimen biner (positif dan negatif) yang diambil dari Hugging Face. Tahap <i>pre-processing</i> meliputi penghapusan <i>stop words</i>, tokenisasi, dan normalisasi teks. Model BERT (<i>bert-base-uncased</i>) di-<i>fine-tuning</i> menggunakan 3 <i>epoch</i> dengan konfigurasi <i>batch size</i> 16, <i>weight decay</i> 0.01, dan <i>learning rate warmup</i>. Hasil evaluasi menunjukkan akurasi 94% pada data <i>testing</i>, dengan F1-score seimbang (0.94) untuk kedua kelas. <i>Confusion matrix</i> mengungkapkan performa model lebih baik dalam mengidentifikasi sentimen negatif (<i>recall</i> 95%) dibanding positif (<i>recall</i> 92%). Penelitian ini membuktikan efektivitas BERT untuk analisis sentimen berbasis teks dalam konteks <i>e-commerce</i> Indonesia, dengan rekomendasi pengembangan melalui perbanyakan sampel data dan eksperimen varian BERT lainnya untuk peningkatan akurasi</p>
<p><b>Keywords:</b> <i>E-Commerce</i> Indonesia NLP <i>Sentiment Analysis</i> BERT</p>		
		Copyright © 2025. All rights reserved.

## I. Introduction

Di Indonesia, perkembangan *e-commerce* mengalami percepatan yang signifikan, terutama selama masa pandemi COVID-19. Pandemi ini menyebabkan perubahan perilaku konsumen yang sebelumnya lebih banyak melakukan transaksi secara tunai dan tatap muka, kini beralih ke transaksi digital melalui platform daring. Pertumbuhan *e-commerce* ini juga didukung oleh kemudahan akses teknologi, berbagai program promosi menarik, pengaruh sosial, iklan yang masif, serta tren yang berkembang di masyarakat [1]. Berdasarkan laporan e-Conomy SEA 2024 oleh Google, Temasek, serta Bain & Company, *Gross Merchandise Value* (GMV) sektor digital Indonesia diproyeksikan mencapai USD 90 miliar pada tahun 2024, mengalami pertumbuhan sebesar 13% dibandingkan tahun sebelumnya [2]. Prestasi ini tidak hanya mencerminkan ketahanan ekonomi digital nasional, tetapi juga menegaskan posisi Indonesia sebagai pasar digital terbesar di kawasan Asia Tenggara. Menurut laporan Kementerian Perdagangan tentang perdagangan digital Indonesia periode 2023, jumlah pengguna *e-commerce* menunjukkan pertumbuhan signifikan dari tahun 2020 hingga 2023 dengan 58.63 juta pengguna, dan diperkirakan akan terus meningkat hingga mencapai 99.1 juta pengguna pada tahun 2029. Pertumbuhan ini didorong oleh dominasi beberapa platform utama seperti Shopee, Tokopedia, Lazada, Bukalapak, Blibli, dan Zalora, yang secara kolektif menguasai sebagian besar pasar *e-commerce* nasional. Laporan Kementerian Perdagangan tersebut juga mengungkapkan bahwa Tokopedia dan Shopee menjadi platform paling banyak diakses, dengan masing-masing mencatat 158.35 juta dan 131.3 juta kunjungan pada kuartal II-2022.

Dengan jutaan transaksi yang terjadi setiap hari di berbagai platform *e-commerce* di Indonesia, ulasan produk menjadi komponen penting dalam pengambilan keputusan konsumen. Setiap pembelian biasanya disertai ulasan yang mencerminkan pengalaman pengguna, mulai dari kualitas produk, ketepatan pengiriman, hingga pelayanan penjual. Seluruh ulasan tersebut tidak hanya membantu calon pembeli membuat keputusan yang lebih baik, tetapi juga dapat memberikan

informasi mengenai preferensi serta tingkat kepuasan konsumen bagi pelaku bisnis dan platform *e-commerce* itu sendiri. *Volume* data ulasan yang masif dan terus bertambah setiap harinya mengakibatkan analisis manual menjadi tidak efisien dan kurang praktis untuk diterapkan, sehingga dibutuhkan metode otomatis seperti analisis sentimen.

Analisis sentimen adalah cabang *Natural Language Processing* (NLP) yang digunakan untuk mengkategorikan emosi dalam teks [4]. Analisis sentimen telah mengalami evolusi signifikan dari pendekatan berbasis model *machine learning* konvensional menuju teknik berbasis *deep learning*. Pendekatan berbasis model *machine learning* konvensional seperti Naive Bayes, *Support Vector Machines* (SVM) dan *decision tree* bergantung pada ekstraksi fitur manual, yang meskipun efisien secara komputasi, memiliki keterbatasan dalam menangkap konteks dan nuansa bahasa [5]. *Deep learning* memungkinkan klasifikasi sentimen yang lebih akurat dan bernuansa, penemuan model bahasa *pre-trained* seperti *Bidirectional Encoder Representations from Transformers* (BERT) dan *Generative Pretrained Transformer* (GPT) semakin memajukan analisis sentimen. Model-model ini dilatih pada data teks dalam jumlah besar, mampu memahami konteks, sarkasme, dan fitur bahasa kompleks lainnya yang sering menjadi tantangan bagi model konvensional.

Penelitian ini bertujuan untuk melakukan analisis sentimen positif dan negatif pada ulasan produk di platform *e-commerce* Indonesia menggunakan BERT. Model BERT yang digunakan adalah versi *pre-trained* (*bert-base-uncased*) yang dikembangkan oleh Google, dengan penyesuaian (*fine-tuning*) khusus untuk dataset ulasan dalam bahasa Indonesia. Dataset yang digunakan diambil dari Hugging Face (<https://huggingface.co/datasets/dipawidia/ecommerce-product-reviews-sentiment>), yang terdiri dari 11606 ulasan produk dari berbagai merek dan produk Indonesia di platform *e-commerce* seperti Shopee, Tokopedia, Lazada, Bukalapak, Blibli, dan Zalora.

## II. Literature Review

### 2.1. *Natural Language Processing* (NLP)

*Natural Language Processing* (NLP) adalah bidang dalam *Artificial Intelligence* (AI) dan linguistik yang bertujuan untuk membuat komputer mampu memahami pernyataan atau kalimat dalam bahasa manusia. NLP dapat diklasifikasikan ke dalam dua bagian, yaitu *Natural Language Understanding* (NLU) and *Natural Language Generation* (NLG). NLU merupakan ilmu yang mempelajari bahasa, mencakup berbagai aspek penting, seperti *phonology* yang mempelajari bunyi bahasa, *morphology* tentang pembentukan kata, *syntax* yang mengkaji struktur kalimat, *semantic* yaitu makna bahasa, dan *pragmatic* yang berfokus pada pemahaman konteks penggunaan bahasa [6]. Selain NLU, terdapat pula *Natural Language Generation* (NLG) yang merupakan proses pembuatan frasa, kalimat, maupun paragraf yang koheren dan bermakna dari suatu representasi internal. NLG berperan penting dalam mengubah data terstruktur menjadi bahasa alami yang dapat dipahami manusia.

NLP digunakan secara luas dalam berbagai bidang. Dalam dunia bisnis, teknologi ini dimanfaatkan untuk *chatbot* layanan pelanggan, analisis umpan balik konsumen, serta penyaringan otomatis *email spam*. Pada bidang kesehatan, NLP membantu ekstraksi informasi dari rekam medis elektronik dan literatur medis untuk mendukung diagnosis dan penelitian. Sementara itu, di media sosial, algoritma NLP digunakan untuk memantau tren percakapan, mendeteksi konten berbahaya, atau bahkan menghasilkan ringkasan otomatis dari diskusi panjang.

### 2.2. *Sentiment Analysis*

Analisis sentimen adalah cabang *Natural Language Processing* (NLP) yang digunakan untuk mengkategorikan emosi dalam teks [4]. Analisis sentimen telah mengalami evolusi signifikan dari *rule-based approaches*, pendekatan berbasis model *machine learning* konvensional hingga teknik berbasis *deep learning*. Pada era 1960–1970-an, pendekatan berbasis aturan (*rule-based approaches*) menjadi fondasi analisis teks, di mana sistem bergantung pada *dictionaries* sentimen yang telah ditentukan. Namun, metode ini memiliki keterbatasan dalam menangani konteks dan nuansa bahasa. Memasuki 1980-an, teknik berbasis leksikon mulai digunakan, meskipun masih kurang memahami konteks secara mendalam. Selanjutnya pada 1990-an, analisis sentimen berkembang dengan pendekatan berbasis *keywords* dan *dictionaries* (WordNet), tetapi tantangan seperti sarkasme dan ironi masih sulit diatasi.

Perkembangan besar terjadi di tahun 2000-an dengan hadirnya metode *machine learning* seperti SVM dan Naive Bayes, yang memanfaatkan dataset berlabel dan fitur statistik seperti TF-IDF. Namun, pendekatan ini masih terbatas dalam menangkap hubungan semantik antar kata. Revolusi *deep learning* di tahun 2010-an membawa terobosan dengan model RNN, LSTM, dan Word2Vec, yang mampu memahami konteks berurutan. Momen penting selanjutnya dalam analisis sentimen terjadi dengan diperkenalkannya arsitektur *transformer* beserta mekanisme *self-attention*-nya [8]. Berbeda dengan RNN yang memproses data secara berurutan (linear), *transformer* memungkinkan model untuk memproses seluruh kalimat atau dokumen secara paralel sambil mempertimbangkan hubungan antar semua kata secara bersamaan. Mekanisme *attention* ini memungkinkan model untuk menimbang pentingnya setiap kata relatif terhadap kata lainnya, sehingga mampu menangkap dependensi jarak jauh dan makna kontekstual dengan lebih efektif. Kemunculan model berbasis *transformer* seperti BERT dan GPT menandai era baru bagi analisis sentimen dan NLP secara keseluruhan.

### 2.3. Bidirectional Encoder Representations from Transformers (BERT)

*Bidirectional Encoder Representations from Transformers* (BERT) adalah model berbasis arsitektur *transformer* yang digunakan dalam NLP, menjawab pertanyaan, dan analisis sentimen. Dikembangkan oleh peneliti Google pada tahun 2018, BERT merupakan model *pre-trained* yang unik karena bersifat *bidirectional* (dua arah), memproses teks dari kiri ke kanan dan sebaliknya. Model ini dilatih menggunakan *corpus* teks yang sangat besar, termasuk Wikipedia dan kumpulan buku. Berbeda dengan model konvensional yang menghasilkan embedding kata tunggal tanpa mempertimbangkan konteks, BERT bersifat kontekstual. Seperti ditunjukkan oleh Dong *et al.* (2013) [9], BERT mempertimbangkan konteks sebelum dan sesudah kata untuk menghasilkan representasi yang lebih akurat. BERT dilakukan melalui dua pendekatan utama: (1) *Masked Language Prediction*, yaitu sebagian kata dalam teks disembunyikan (*masked*), lalu BERT memprediksi kata tersebut berdasarkan konteks kata-kata di sekitarnya. (2) *Next Sentence Prediction*, BERT menerima dua kalimat sebagai *input* dan menentukan apakah kalimat kedua merupakan kelanjutan logis dari kalimat pertama atau tidak.

### 2.4. Related Research

Salah satu penelitian yang cukup signifikan dilakukan pada ulasan produk iPhone di platform Amazon. Dalam studi ini, Setyawan *et al.* (2024) [10] memanfaatkan model BERT untuk mengklasifikasikan sentimen dari ribuan ulasan yang diambil dari Kaggle. Dengan melakukan *preprocessing* seperti tokenisasi dan pelabelan sentimen, model BERT berhasil mencapai akurasi hingga 93.9% dan F1-score yang seimbang. Hasil ini menunjukkan keunggulan BERT dalam menangkap konteks kalimat yang kompleks pada ulasan produk. Selvakumar *et al.* (2022) [11] menggunakan model BERT untuk klasifikasi sentimen ulasan pengguna pada *dataset* IMDB *Movie Review* dan Amazon *Fine Food Review*, menunjukkan bahwa BERT lebih efektif dalam menangkap hubungan sentimen dibandingkan model *machine learning* dan *deep learning* dasar. Akurasi model BERT mencapai 94%, sedangkan akurasi dari model lain umumnya berkisar pada rentang  $\pm 70\%$  hingga  $\pm 85\%$  (kecuali model *Lexicon with Multilayer Perceptron* yang memiliki akurasi sebesar 91.9%).

Penelitian serupa juga dilakukan pada platform *e-commerce* di Indonesia. Salah satunya adalah analisis ulasan aplikasi Bilibili dengan menggunakan BERT yang digabungkan dengan metode *clustering* KMeans untuk pelabelan aspek *usable*, *valuable*, dan *credible* oleh Rahman *et al.* (2025) [12]. Hasil penelitian ini sangat mengesankan, dengan akurasi mencapai 99% dan F1-score sebesar 98.9%. Studi ini membuktikan bahwa BERT juga efektif digunakan pada data berbahasa Indonesia, terutama setelah dilakukan *fine-tuning* pada dataset yang relevan. Selain itu, penelitian pada aplikasi Segari, sebuah platform *quick commerce* di Indonesia, juga mengaplikasikan BERT untuk analisis sentimen ulasan dari *Google Play Store* [13]. Dengan teknik *web scraping* untuk pengumpulan data, model BERT yang digunakan mampu mencapai akurasi 89% dan presisi 91% untuk sentimen positif. Keberhasilan ini memperkuat temuan bahwa BERT sangat adaptif terhadap berbagai jenis data dan bahasa, serta dapat diandalkan untuk meningkatkan kualitas analisis sentimen di berbagai platform *e-commerce*.

### III. Data and Methods

#### A. Data

Dataset diambil dari Hugging Face (<https://huggingface.co/datasets/dipawidia/ecommerce-product-reviews-sentiment>), yang terdiri dari 11606 ulasan produk dari berbagai merek dan produk Indonesia di platform *e-commerce* seperti Shopee, Tokopedia, Lazada, Bukalapak, Blibli, dan Zalora.

#### B. Methods

##### 1. Data Preprocessing

Dataset memiliki 3 kolom, yaitu review yang berisi ulasan produk dalam bahasa Indonesia, translate berisi ulasan yang telah diterjemahkan ke bahasa Inggris, dan sentimen berisi label ulasan dengan 1 sebagai indikator sentimen positif dan 0 sentimen negatif. Penelitian ini akan menggunakan kolom translate dan sentimen, penggunaan ulasan dalam bahasa Inggris dipilih karena penulis ingin menggunakan model BERT *pre-trained* yang dilatih menggunakan *dictionaries* bahasa Inggris.

- Penghapusan *missing values* dan duplikasi pada data.
- Tokenisasi: Memecah teks menjadi kata-kata individu.
- Mengubah format huruf menjadi kecil.
- Menghapus angka, tanda baca, dan emoji. Lalu, penghapusan *English stop words* (kata dalam bahasa Inggris yang sering terdapat dalam teks, tetapi kurang bermakna).
- Normalisasi: Penggabungan kembali list kata yang sudah diproses menjadi satu *string* teks.

##### 2. BERT Algorithm

- Pendefinisian variabel X dan y, dengan X adalah kolom ulasan dan y adalah kolom label sentimen.
- *Split* data menjadi data *training*, validasi dan *testing*. Pertama, data dibagi menjadi 80% untuk *training* dan 20% data untuk validasi. Kemudian, data validasi dibagi lagi menjadi 50% untuk validasi dan 50% untuk *testing*.
- Tokenisasi X *training* dan *testing* menggunakan BERT *pre-trained* (*bert-base-uncased*) *tokenizer*. Tokenisasi mengonversi teks menjadi format yang kompatibel dengan BERT, dengan menambahkan *padding* agar semua *sequence* memiliki panjang seragam, memotong teks yang melebihi 128 token, dan mengembalikan hasilnya dalam bentuk *dictionary* berisi 'input\_ids', 'attention\_mask', dan 'token\_type\_ids'.
- Membuat PyTorch Dataset untuk menangani data teks yang sudah ditokenisasi dengan BERT. *Class* ini menerima *output* dari BERT *tokenizer* dan label klasifikasi, lalu mengemasnya dalam format yang kompatibel dengan PyTorch. Setiap sampel diakses via '*\_\_getitem\_\_*' yang mengonversi data ke tensor, dan *dataset* ini bisa langsung dipakai dengan 'DataLoader' untuk *training model*. Proses ini untuk menghubungkan antara tokenizer BERT dan model PyTorch, sekaligus mengelola data secara efisien selama pelatihan.
- Inisialisasi model BERT: Memuat model BERT yang sudah dilatih sebelumnya (*bert-base-uncased*) khusus untuk tugas klasifikasi teks menggunakan 'BertForSequenceClassification'. Parameter 'num\_labels=2' mengatur model untuk klasifikasi biner (sentimen positif/negatif), sementara 'output\_attentions=False' dan 'output\_hidden\_states=False' menonaktifkan *output attention weights* dan *hidden states* untuk mengoptimalkan penggunaan memori dan kecepatan pelatihan, karena keduanya tidak diperlukan untuk tugas klasifikasi dasar. Model ini secara otomatis akan menambahkan sebuah *classification head* di atas *pre-trained* BERT.
- Konfigurasi *training arguments*: Objek 'TrainingArguments' digunakan untuk mengatur seluruh parameter pelatihan, 'output\_dir='./results' menentukan direktori penyimpanan output, 'num\_train\_epochs=3' mengatur jumlah *epoch* pelatihan, 'per\_device\_train\_batch\_size=16' dan 'per\_device\_eval\_batch\_size=64' mengatur

ukuran *batch* untuk *training* dan evaluasi. ‘warmup\_steps=500’ mengontrol fase *warm up learning rate scheduler*, ‘weight\_decay=0.01’ menerapkan regularisasi L2, ‘logging\_dir='./logs’ menyimpan *log training*, dan ‘logging\_steps=10’ menentukan frekuensi *logging*. ‘load\_best\_model\_at\_end=True’ dan ‘metric\_for\_best\_model='accuracy'’ memungkinkan penyimpanan model terbaik berdasarkan akurasi, sementara ‘eval\_strategy’ dan ‘save\_strategy’ yang di-set ke ‘epoch’ membuat evaluasi dan penyimpanan model dilakukan tiap akhir *epoch*.

- Pendefinisian evaluasi metrik: Fungsi ‘compute\_metrics’ mendefinisikan cara menghitung metrik evaluasi selama pelatihan. Parameter ‘pred’ menerima prediksi model dan label sebenarnya. ‘np.argmax(pred.predictions, axis=1)’ mengubah *output* model menjadi prediksi kelas, sementara ‘precision\_recall\_fscore\_support’ menghitung *precision*, *recall*, dan *F1-score* dengan metode ‘weighted’ untuk menangani ketidakseimbangan kelas. ‘accuracy\_score’ menghitung akurasi klasifikasi. Hasilnya dikembalikan dalam bentuk *dictionary* yang berisi semua metrik tersebut untuk memantau performa model.
- Inisialisasi trainer: Class ‘Trainer’ mengintegrasikan semua komponen yang telah disiapkan. Parameter ‘model’ menerima model BERT yang telah diinisialisasi, ‘args’ mengambil konfigurasi pelatihan dari ‘TrainingArguments’, ‘train\_dataset’ dan ‘eval\_dataset’ menentukan data *training* dan validasi, serta ‘compute\_metrics’ menentukan fungsi evaluasi yang akan digunakan. Trainer ini akan secara otomatis menangani seluruh proses pelatihan, termasuk *forward pass*, *backward pass*, evaluasi, *logging*, dan penyimpanan model berdasarkan konfigurasi yang telah ditetapkan.
- Lakukan proses *fine-tuning* model BERT berdasarkan konfigurasi yang telah ditetapkan sebelumnya.

### 3. Evaluation Metrics

Mengevaluasi performa model BERT yang telah melalui proses *fine-tuning* pada data *testing*. Pertama, ‘trainer.predict()’ menghasilkan prediksi dalam bentuk *logits*, yang kemudian dikonversi ke label kelas menggunakan ‘np.argmax()’. Lalu, akan ditampilkan laporan klasifikasi yang berisi metrik utama seperti *precision*, *recall*, dan *F1-score* untuk setiap kelas (Negatif dan Positif). Selanjutnya, *confusion matrix* divisualisasikan dengan *heatmap* untuk membandingkan prediksi dan label asli. Terakhir, akurasi keseluruhan dihitung menggunakan ‘accuracy\_score()’. Hasilnya memberikan gambaran komprehensif tentang kinerja model dalam mengklasifikasikan sentimen teks.

## IV. Results and Discussions

Pada data yang diambil dari Hugging Face dilakukan pengecekan *missing values* dan duplikasi. Proses ini menyisakan 11585 observasi dari yang sebelumnya berjumlah 11606. Kolom *translate* di-*rename* menjadi *review*. Setelah itu, dilakukan *pre-processing* pada teks *review*.

Table 1. Perbandingan Teks Sebelum dan Sesudah *Pre-Processing*

Review Before Pre-Processing	Review After Pre-Processing
sound and mic satisfied. I only ordered the one that was on and I made sure that it clicked again, but the one that came didn't turn on	sound mic satisfied ordered one made sure clicked one came turn
good. hopefully durable. Thanks	good hopefully durable thanks
the condition of the box is torn, please pay attention again	condition box torn please pay attention
product quality is good. affordable prices and very fast delivery. I've already ordered here, it's okay, but unfortunately, if you make a gift, you can't bring it with a greeting card or wrap it in gift paper :(	product quality good affordable prices fast delivery already ordered okay unfortunately make gift ca bring greeting card wrap gift paper
overall is good, the buttons all work, but the packaging is lacking because it's dented and torn a little	overall good buttons work packaging lacking dented torn little

Dapat dilihat terdapat perbedaan signifikan antara teks sebelum *pre-processing* dan sesudah. Teks hasil *pre-processing* sudah dalam bentuk huruf kecil semua, tidak lagi terdapat tanda baca dan emoji, *stop words* seperti *is*, *the*, *I*, *but*, *in*, dll juga sudah dihapuskan untuk mengurangi *noise*. Langkah-langkah ini secara kolektif meningkatkan efisiensi ekstraksi fitur untuk BERT.

Data ini telah memiliki jumlah observasi per kelas yang seimbang sehingga tidak memerlukan *resampling*. Pembagian data menjadi *training*, validasi, dan *testing* menghasilkan data *training* berukuran 9268, validasi berukuran 1158, dan *testing* berukuran 1159. Kemudian dilakukan *fine-tuning* model BERT, didapatkan hasil sebagai berikut.

Table 2. Hasil Metrik setiap *Epoch* pada Proses *Fine-Tuning* model BERT

Epoch	Training Loss	Validation Loss	Acuracy	F1	Precision	Recall
1	0.186100	0.211327	0.927461	0.927461	0.927466	0.92746
2	0.159800	0.205366	0.937824	0.937757	0.939718	0.937824
3	0.019000	0.238807	0.943869	0.943859	0.944167	0.943869

Berdasarkan tabel di atas dapat dilihat bahwa model BERT menunjukkan peningkatan kinerja yang konsisten, dengan akurasi naik dari 92.75% (Epoch 1) menjadi 94.39% (Epoch 3), didukung oleh peningkatan *F1-score*, *presicion*, dan *recall*. *Training loss* turun dari 0.186 (Epoch 1) ke 0.019 (Epoch 3), membuktikan model berhasil mempelajari pola data *training*. Namun, pada *validation loss* terdapat sedikit kenaikan di Epoch 3 (0.239) yang mungkin mengindikasikan tanda *overfitting* ringan, meski tidak signifikan karena akurasi validasi tetap naik.

Terakhir, dilakukan evaluasi terakhir dengan menggunakan data *testing*. Dihasilkan *classification report* dan *confusion matrix* sebagai berikut.

Classification Report:					
	precision	recall	f1-score	support	
Negatif	0.93	0.95	0.94	579	
Positif	0.95	0.92	0.94	580	
accuracy			0.94	1159	
macro avg	0.94	0.94	0.94	1159	
weighted avg	0.94	0.94	0.94	1159	

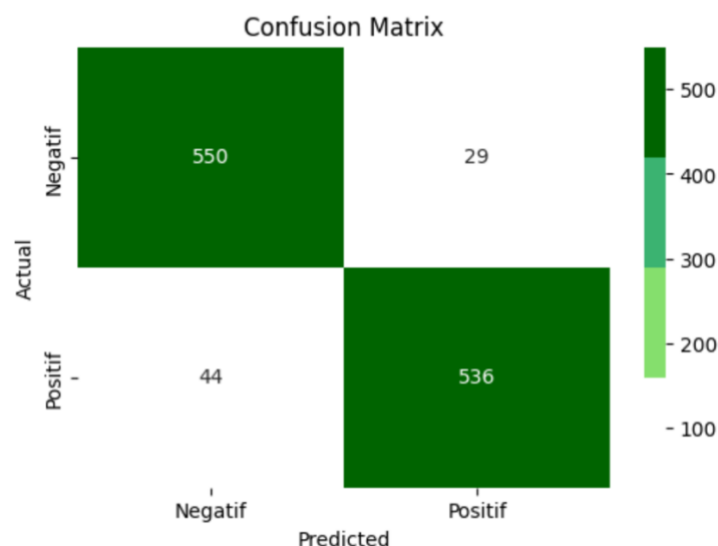
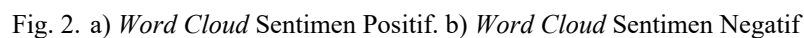


Fig. 1. *Classification Report* dan *Confusion Matrix* untuk Data *Testing*

Berdasarkan gambar di atas model menunjukkan performa yang sangat seimbang dengan akurasi keseluruhan 94% pada data *testing*. Hal ini juga terlihat dari nilai *precision*, *recall*, dan *F1-score* yang konsisten di kisaran 0.93-0.95 untuk kedua kelas (Negatif dan Positif). Selain itu, gambar 1 mengungkapkkan bahwa model sedikit lebih baik dalam mengidentifikasi ulasan Negatif (*recall* 95%)



Sebagai tambahan penelitian ini juga menampilkan distribusi dari kata yang muncul pada ulasan dalam bentuk *word cloud*.



## V. Conclusions

## References

- Nadira Eka Rahmaharva (Analisis Sentimen pada Ulasan Produk....)*

- [8] M. Jiang, J. Wu, X. Shi, and M. Zhang, "Transformer based memory network for sentiment analysis of web comments," *IEEE Access*, vol. 7, pp. 179942–179953, 2019.
- [9] R. Dong, M. Schaal, M. P. O'Mahony, and B. Smyth, "Topic Extraction from Online Reviews for Classification and Recommendation," in *Proc. IJCAI Int. Joint Conf. Artif. Intell.*, Beijing, China, 2013, pp. 1310–1316.
- [10] A. F. Setyawan, A. D. P. Ariyanto, F. K. Fikriah, and R. I. Nugraha, "Analisis sentimen ulasan iPhone di Amazon menggunakan model deep learning BERT berbasis transformer," *Elkom: Jurnal Elektronika dan Komputer*, 2024.
- [11] B. Selvakumar and B. Lakshmanan, "Sentimental analysis on user's reviews using BERT," *Materials Today: Proceedings*, vol. 62, part 7, pp. 4931-4935, 2022. doi: 10.1016/j.matpr.2022.03.678.
- [12] R. D. Rahman, N. Y. Setiawan, and F. A. Bachtar, "Analisis Sentimen Pengguna Aplikasi Mobile Berbasis Review Pada Platform Bibli Menggunakan Metode Bidirectional Encoder Representations from Transformers (BERT)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, pp. x-x, Apr. 2025.
- [13] Andriani Marshanda Putri, Widya Khafa Nofa, and Dewi Anggraini Puspa Hapsari, "PENERAPAN METODE BERT UNTUK ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SEGARI DI GOOGLE PLAY STORE", *JUIT*, vol. 4, no. 1, pp. 89–104, Jan. 2025.