

Perbandingan Algoritma *Naive Bayes* dan *Bert* untuk Analisis Sentimen Ulasan Produk *Shopee* Berdasarkan Rating dan Atribut Produk (Warna/Kategori)

Petrus Maxmiliano^{1*}, Yosefina Finsensia Riti², Isaac Yeremia Nugroho³, Claudio Erlisto Candra Juniarto⁴

^{1,2,3,4}Ilmu Informatika, Universitas Katolik Darma Cendika, Indonesia

Email: ¹petrus.maxmiliano@student.ukdc.ac.id, ²yosefina.riti@ukdc.ac.id, ³isaac.nugroho@student.ukdc.ac.id,

⁴claudio.juniarto@student.ukdc.ac.id

Email Penulis Korespondensi: ²yosefina.riti@ukdc.ac.id

Abstrak—Perkembangan E-commerce yang berkembang pesat, khususnya melalui platform *Shopee*, telah menghasilkan banyak ulasan produk dari pengguna yang dapat dimanfaatkan untuk mengetahui persepsi konsumen terhadap suatu produk. Penelitian ini bertujuan untuk membandingkan performa algoritma *Naive Bayes* dan BERT dalam analisis sentimen ulasan produk *Shopee* berdasarkan rating dan atribut produk seperti warna dan kategori. Penelitian ini menggunakan dataset berjumlah 2.300 ulasan yang telah melalui tahap pelabelan dan pra-pemrosesan (*case Folding*, Tokenisasi) serta pemodelan menggunakan algoritma BERT dan *Naive Bayes*. Hasil evaluasi menunjukkan bahwa model BERT memiliki performa yang lebih unggul dibandingkan *Naive Bayes*, dengan akurasi sebesar 94,9% dan F1-score yang lebih merata di setiap kelas. BERT juga mampu menangkap konteks ulasan yang lebih kompleks dan menghasilkan klasifikasi sentimen yang lebih konsisten. Sementara itu, *Naive Bayes* menunjukkan performa yang cukup baik namun cenderung kurang stabil dalam mendeteksi sentimen berdasarkan atribut tertentu. Untuk mengevaluasi performa model, digunakan metode evaluasi berbasis confusion matrix dengan metrik akurasi, presisi, recall, dan F1-score. Sedangkan proses implementasi nya menggunakan bantuan tools seperti python, library scikit-learn, dan Hugging Face Transformers. Berdasarkan hasil evaluasi diatas, BERT direkomendasikan untuk analisis sentimen ulasan e-commerce karena konsistensinya dalam menangkap konteks kalimat yang kompleks. Untuk penelitian selanjutnya disarankan menambahkan atribut lain seperti harga dan waktu pengiriman, memperluas dataset, serta mengeksplorasi algoritma lain seperti LSTM, dan Random Forest guna meningkatkan akurasi dan relevansi hasil.

Kata Kunci: E-commerce, *Shopee*, Analisis Sentimen, NLP, BERT, *Naive Bayes*, Ulasan produk, Klasifikasi

Abstract—The rapid development of e-commerce, particularly through the *Shopee* platform, has generated numerous user product reviews that can be used to determine consumer perceptions of a product. This study aims to compare the performance of the *Naive Bayes* and BERT algorithms in sentiment analysis of *Shopee* product reviews based on ratings and product attributes such as color and category. This study used a dataset of 2.300 reviews that had gone through labeling and pre-processing stages (case folding, tokenization) and modeling using the BERT and *Naive Bayes*. The evaluation results showed that the BERT model has superior performance compared to *Naive Bayes*, with an accuracy of 94.9% and a more even F1-score across each class. BERT was also able to capture more complex review contexts and produce more consistent sentiment classifications. Meanwhile, *Naive Bayes* shows quite good performance but tended to be less stable in detecting sentiment based on certain attributes. To evaluate model performance, a confusion matrix-based evaluation method was used with accuracy, precision, recall, and F1-score metrics. Meanwhile, the implementation process used tools such as Python, the scikit-learn library, and Hugging Face Transformers. Based on the evaluation results above, BERT is recommended for sentiment analysis of e-commerce reviews due to its consistency in capturing complex sentence context. Future research suggests adding other attributes such as price and delivery time, expanding the dataset, and exploring other algorithms such as LSTM and Random Forest to improve the accuracy and relevance of the results.

Keywords: E-commerce, *Shopee*, Analysis sentiment, NLP, BERT, *Naive Bayes*, review product, classification

1. PENDAHULUAN

Perdagangan elektronik (*e-commerce*) merupakan seluruh aktivitas jual beli yang dilakukan secara daring antara penjual dan pembeli melalui jaringan internet maupun perangkat elektronik lainnya. Di era transformasi digital saat ini, pelaku bisnis dari berbagai skala baik dari perusahaan besar maupun usaha kecil dan menengah semakin banyak yang beralih atau mengembangkan usahanya melalui platform digital, khususnya dalam bentuk *e-commerce* [1]. *Shopee* adalah salah satu platform e-commerce yang sangat populer di Indonesia dan memiliki daya tarik yang sangat tinggi bagi masyarakat. Berdasarkan data yang dirilis oleh iPrice, *Shopee* menempati posisi teratas dengan jumlah kunjungan tertinggi di Indonesia, yakni mencapai 96 juta kunjungan pada kuartal ketiga dan meningkat signifikan menjadi 129 juta pada kuartal keempat tahun 2020. Di kawasan Asia Tenggara, *Shopee* menjadi salah satu platform *e-commerce* terkemuka dengan total kunjungan sekitar 281 juta. Tingginya minat masyarakat terhadap *Shopee* menyebabkan banyaknya ulasan produk yang diberikan oleh pembeli. Ulasan produk sangat bermanfaat bagi konsumen karena dapat memberikan informasi mengenai kualitas barang. Melalui pengalaman yang diberikan oleh pengguna lain, calon pembeli dapat mempertimbangkan keputusan sebelum melakukan transaksi.[2].

Analisis sentimen merupakan proses dalam ilmu komputasi yang bertujuan untuk memahami, mengekstraksi, serta memproses data dalam bentuk teks secara otomatis guna memperoleh informasi dari opini atau perilaku individu.



Pendekatan ini sangat penting untuk mengetahui tanggapan pengguna terhadap suatu produk dengan cara menganalisis ulasan pengguna, sehingga emosi atau persepsi mereka dapat diidentifikasi secara lebih mendalam [3]. Penelitian mengenai analisis sentimen menjadi salah satu topik yang menarik dalam bidang data mining dan pemrosesan bahasa alami (*Natural Language Processing/NLP*), Karena dapat memberikan wawasan penting bagi pengambilan keputusan bisnis berbasis data. Beberapa penelitian sebelumnya lebih banyak berfokus pada penggunaan algoritma klasik seperti atau *Support Vector Machine* (SVM) yang meskipun cukup efektif seringkali kurang mampu menangkap konteks bahasa yang lebih kompleks.

Kesenjangan penelitian terlihat pada keterbatasan metode tersebut dalam memahami makna mendalam dari suatu ulasan, terutama pada kalimat dengan struktur bahasa yang bervariasi. Oleh karena itu, penelitian ini berkontribusi dengan menerapkan pendekatan berbasis *Bidirectional Encoder Representations From Transformers* (BERT), yang mampu memproses konteks kalimat secara lebih mendalam dibandingkan metode konvensional. Dengan demikian, penelitian ini diharapkan dapat memberikan hasil klasifikasi sentimen yang lebih akurat dan relevan dalam konteks e-commerce.

Berbagai penelitian terdahulu telah mengkaji analisis sentimen menggunakan beragam algoritma. [4] melakukan analisis sentimen terhadap opini publik mengenai Permendikbud No.30 dengan menggunakan metode *Naive Bayes* dan LSTM, hasilnya menunjukkan bahwa LSTM memiliki akurasi yang sedikit lebih tinggi (77%) dibandingkan *Naive Bayes* (76%). penelitian lain, oleh [5] membandingkan performa *Naive Bayes* dan KNN dalam mengklasifikasikan opini masyarakat terhadap isu COVID-19 di Twitter, dan menemukan bahwa *Naive Bayes* memberikan hasil klasifikasi yang lebih akurat (63,21%) dibandingkan KNN (58,10%). Hal ini menunjukkan bahwa algoritma sederhana seperti *Naive Bayes* masih kompetitif untuk data teks dengan karakteristik tertentu.

Penelitian lain oleh, [6] melakukan klasifikasi sentimen terhadap opini masyarakat Indonesia mengenai konten *deepfake* menggunakan model *IndoBERT* melalui pendekatan semi-supervised learning, dan mencapai akurasi sebesar 82% serta nilai F1 sebesar 54,3%. Selain itu, metode *Classifier* dalam penelitian tersebut juga menghasilkan recall sebesar 51,34% dan precision sebesar 61%. Hal ini memperlihatkan keunggulan model berbasis Transformer seperti BERT dibandingkan algoritma klasik. Di sisi lain, [7] menganalisis sentimen pengguna aplikasi Segari menggunakan metode BERT dengan proporsi data untuk pelatihan dan pengujian sebesar 70:30, dan menghasilkan akurasi sebesar 89%. Nilai precision untuk masing-masing kategori juga cukup tinggi: positif 91%, negatif 83%, dan netral 69%. Hal ini menegaskan konsistensi performa BERT dalam menangani data ulasan yang kompleks.

[8] menerapkan algoritma dalam analisis sentimen ulasan pengguna dengan dua kategori (positif dan negatif) menggunakan pendekatan NLP, dan mendapatkan akurasi sebesar 88%. Penelitian lain oleh [9] menunjukkan bahwa untuk label negatif, precision mencapai 0,87, recall 1,00, dan F1-score sebesar 0,93. Sementara itu untuk label positif, precision mencapai 1,00 tetapi recall hanya 0,33, dengan F1-score 0,50 dan mendapatkan akurasi keseluruhan sebesar 88%. Hal ini mengindikasikan adanya keterbatasan dalam menangani data dengan variasi kelas yang tidak seimbang.

Penelitian lain yang dilakukan oleh, [10] mengungkapkan bahwa Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia dengan menggunakan SVM berhasil mencapai akurasi sebesar 96,68%. Sementara itu, [11] membuktikan bahwa penerapan *BidirectionalLSTM* bagi Analisis Sentimen Terhadap Layanan Grab Indonesia mendapatkan akurasi sebesar 91%. Meskipun demikian, *Bidirectional LSTM* membutuhkan dataset yang besar untuk menghindari overfitting serta biaya komputasi yang tinggi.

Hasil studi dari beberapa penelitian juga menunjukkan bahwa pemilihan algoritma sangat berpengaruh pada performa klasifikasi. Misalnya [12] melakukan analisis sentiment ulasan aplikasi Genshin Impact di Google Play Store menggunakan algoritma Random Forest, model tersebut berhasil mencapai akurasi hingga 90,2%. Hal ini menunjukkan bahwa algoritma yang tepat dapat menghasilkan performa yang tinggi pada data ulasan yang kompleks. Disisi lain, [13] penelitian komparatif analisis sentiment pemilu presiden RI 2024, dan menemukan bahwa Multinomial *Naive Bayes* menjadi algoritma dengan performa terbaik, yakni dengan akurasi rata-rata sebesar 79% dibandingkan dengan algoritma yang lain seperti KNN, Decision Tree, dan Stochastic Gradient Descent.

Studi lain, oleh [14] mengenai Ulasan di Aplikasi Threads dan Twitter menggunakan perbandingan antara algoritma *Naive Bayes* dengan akurasi 85.56%, KNN 71.67%, dan Decision Tree 72.78%, pada aplikasi Twitter, sementara itu untuk aplikasi *Threads* mendapatkan akurasi sebesar 66.41%, KNN mencapai sebesar 62.50% , dan Decision Tree sebesar 65.41%. Kesimpulan dari hasil akurasi penelitian ini menunjukkan bahwa lebih unggul meskipun akurasi nya belum mencapai angka sempurna. Namun, Sebagian besar studi tersebut belum mempertimbangkan atribut produk seperti harga, kualitas dan waktu pengiriman sebagai variabel analisis sentiment. Kesenjangan inilah yang diisi oleh penelitian ini dengan mengeksplorasi penggunaan BERT dan *Naive Bayes* pada ulasan e-commerce yang mempertimbangkan atribut produk sehingga hasil analisis lebih akurat dan relevan bagi pengambilan Keputusan bisnis.

Berdasarkan hasil studi tersebut, Penggunaan algoritma *Naive Bayes* dan BERT dalam analisis sentiment ulasan produk *Shopee* didasarkan pada kekuatan dan keunggulan masing-masing pendekatan. *Naive Bayes* dikenal sebagai

algoritma yang sederhana, efisien secara komputasi, dan telah terbukti dapat memberikan hasil yang cukup akurat dalam berbagai penelitian sebelumnya, seperti pada analisis opini public, isu sosial, dan ulasan aplikasi dengan akurasi mencapai 88% dan performa yang konsisten lebih baik dibandingkan algoritma klasik lainnya seperti KNN dan Decision Tree.

Disisi lain, BERT merupakan model berbasis deep learning yang mampu memahami konteks Bahasa secara mendalam dan bidirectional, sangat efektif dalam menangani data teks yang kompleks, termasuk ulasan konsumen dalam Bahasa Indonesia. Studi sebelumnya menunjukkan bahwa BERT, terutama IndoBERT, mampu mencapai akurasi tinggi hingga 89% dengan precision kategori positif mencapai 91% dan menjadikan sangat relevan untuk analisis sentiment berbasis atribut produk dan rating. Oleh karena itu, kombinasi penggunaan *Naïve Bayes* dan BERT memungkinkan perbandingan performa antara pendekatan tradisional dan modern dalam menghasilkan klasifikasi sentiment yang lebih akurat dan kontekstual.

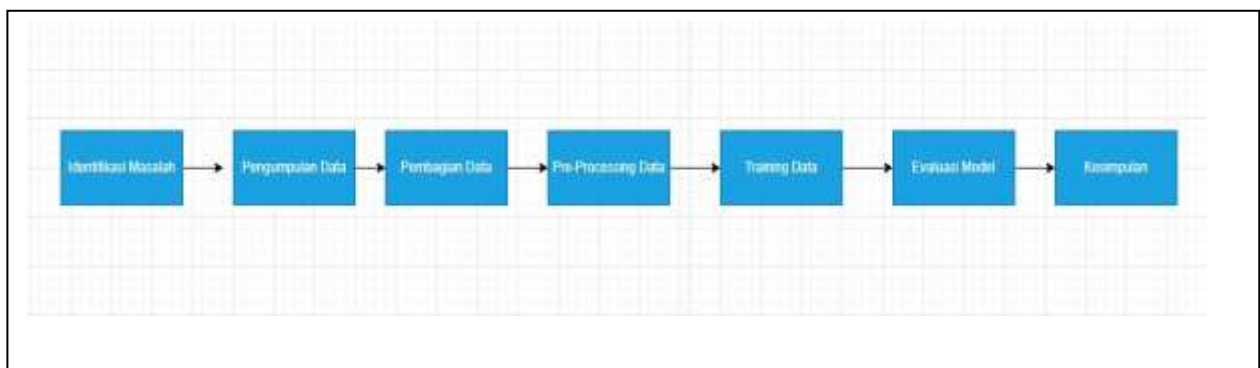
Penelitian ini bertujuan untuk menganalisa sentimen terhadap ulasan produk di platform *Shopee* dengan mempertimbangkan rating pengguna serta atribut produk seperti warna dan kategori. Dua algoritma yaitu *Naive Bayes* dan BERT digunakan untuk membandingkan performa dalam tugas klasifikasi sentimen. Meskipun banyak penelitian telah membandingkan algoritma klasifikasi untuk analisis sentimen, sebagian besar hanya berfokus pada teks ulasan tanpa mempertimbangkan atribut produk seperti warna dan kategori. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan menghadirkan analisis sentimen yang menggabungkan pendekatan tradisional (*Naive Bayes*) dan Modern (BERT) serta memperhatikan atribut produk sebagai variabel tambahan. Diharapkan Hasil penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai preferensi konsumen dan mendukung pengambilan keputusan strategis dalam pengelolaan e-commerce.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode yang digunakan untuk menganalisis sentimen dari ulasan Produk di *Shopee* Berdasarkan Rating dan Kategori melalui pendekatan teknik pemrosesan bahasa alami yang sering disebut Natural Language Processing (NLP) yang dapat ditunjukkan melalui alur diagram pada Gambar 1 berikut. Langkah pertama yakni dengan mengumpulkan data ulasan produk pada aplikasi *Shopee*. Selanjutnya pada tahapan kedua ulasan yang dikumpulkan akan melalui proses pelabelan dan pre-processing. Proses ini mencakup pembersihan teks dari karakter khusus, normalisasi data, serta penghapusan stopwords. Pada tahap ketiga data yang telah diproses akan dibagi menjadi dua bagian yaitu data untuk pelatihan (80%) dan data untuk pengujian (20%). Baru di tahap ke empat akan dilakukan training kedua model yakni *Naive Bayes* dan BERT. Terakhir, pada tahap kelima dilakukan pengujian performa model menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.

Agar peneliti dapat menjalankan ulang eksperimen, kode program dikembangkan menggunakan Python dengan pustaka seperti pandas, sklearn, scikit-learn untuk *Naive Bayes* dengan representasi menggunakan TF-IDF dan Hugging Face Transformers untuk BERT menggunakan framework PyTorch. Struktur dataset terdiri dari beberapa kolom yaitu ulasan (teks ulasan dari pengguna), rating (nilai bintang 1-5), kategori produk, dan warna atau varian. Format input yang digunakan berupa teks ulasan dalam bahasa Indonesia, sedangkan format output berupa label sentimen hasil prediksi dari masing-masing model.



Gambar 1. Diagram Alur Tahapan Penelitian

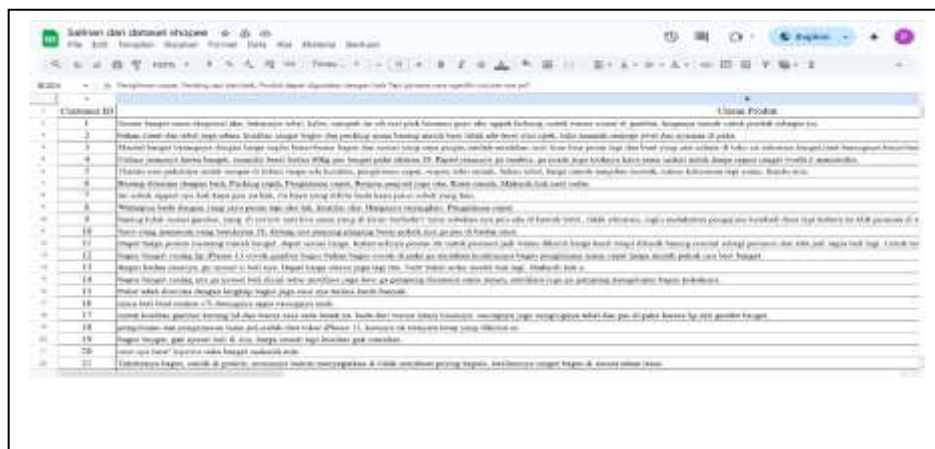
2.2 Pengumpulan Dataset

Dataset yang dipakai dalam penelitian ini dikumpulkan secara manual dari beragam komentar ulasan produk di kalangan masyarakat di Indonesia dengan jumlah total 2.300 data yang diperoleh dari aplikasi *Shopee*

(<https://Shopee.co.id/>). Dataset ini tidak tersedia secara publik, melainkan dikompilasi secara internal saja untuk penelitian ini. Seluruh dataset tersebut disimpan dalam bentuk CSV dengan beberapa kolom yaitu sebagai berikut:

1. Customer ID: ID Unik Pelanggan
2. Ulasan produk: Teks ulasan dari pengguna
3. Rating: Skor penilaian (1-5)
4. Warna varian/ukuran: Informasi atribut warna atau ukuran produk.
5. Kategori: Jenis kategori produk

Setiap ulasan diberikan label sentimen positif atau negatif berdasarkan nilai rating (≥ 4 diberikan label 1 yang berarti positif, ≤ 3 diberikan label 0 yang berarti negatif). Untuk eksperimen, dataset dibagi menjadi rasio 80% data untuk pelatihan, dan 20% untuk data pengujian. Proses split dilakukan menggunakan fungsi `train_test_split` dari pustaka `scikit-learn` dengan parameter `random_state = 42` untuk memastikan konsistensi hasil setiap kali eksperimen dijalankan ulang. Pada gambar 2 di bawah ini merupakan contoh dataset beragam ulasan produk *Shopee* yang digunakan dalam penelitian ini.

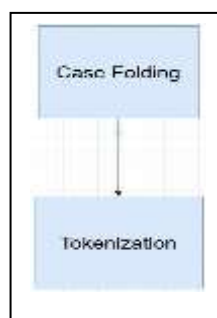


Gambar 2. Dataset di google sheets

2.3 Pre Processing

Metode penelitian yang digunakan tidak hanya berdasarkan tahapan standar NLP, tetapi juga menggunakan pendekatan eksperimental seperti yang dilakukan oleh [15] dimana kombinasi atribut tekstual dan metainformasi seperti rating dan kategori berhasil meningkatkan kinerja klasifikasi sentimen dalam domain produk fashion. Oleh karena itu, dalam penelitian ini, data tidak hanya diproses berdasarkan teks ulasan, tetapi juga menambahkan sebuah atribut tambahan seperti warna dan kategori produk yang relevan untuk memperkaya dalam konteks data.

Selanjutnya, tahap berikutnya adalah preprocessing data. Preprocessing adalah langkah yang diambil untuk menyiapkan data sebelum proses pemodelan. Preprocessing adalah salah satu teknik dalam pengolahan data mining untuk mengubah data mentah menjadi format yang lebih mudah dipahami dikalangan masyarakat Indonesia [16]. Tujuan dari tahapan ini adalah agar data dapat dikelompokkan sehingga mempermudah analisis menggunakan algoritma *Naive Bayes* dan BERT. Pada fase ini data akan mengalami beberapa proses pra-pemrosesan sebelum masuk ke fase validasi dimulai dengan penanganan data, case folding, dan tokenisasi. Langkah-langkah pra-pemrosesan tersebut dapat dilihat pada gambar 3 dibawah ini.



Gambar 3. Diagram Alur Pre Processing



1. *Case Folding* : *Case folding* merupakan langkah dimana semua kata dalam kumpulan data diubah menjadi huruf kecil dengan menggunakan fungsi(*lowercase*)
2. *-Tokenisasi* : Tokenisasi adalah tahap dimana teks yang ada di kalimat dipecah menjadi kata-kata.

Tahapan *preprocessing* mengikuti pendekatan yang digunakan oleh [17] dengan penyesuaian pada atribut produk. Selain itu, [18] juga menerapkan tahapan *case folding* dan tokenisasi dalam perbandingan algoritma *Naive Bayes* dan KNN pada ulasan Google Play Store.

2.4 Metode Analisis

Dalam penelitian ini, penulis menerapkan dua jenis algoritma untuk klasifikasi yaitu *Naive Bayes* dan BERT. Penulis memilih kedua algoritma ini untuk mengetahui mana yang memiliki tingkat akurasi yang paling tinggi. Pemilihan algoritma BERT didasarkan pada hasil penelitian sebelumnya yang menunjukkan kinerja yang optimal. Di samping itu, penulis juga memilih algoritma *Naive Bayes* karena metode ini sangat terkenal dan sering digunakan di kalangan masyarakat Indonesia. Model BERT: Tokenisasi dilakukan menggunakan BertTokenizer, yang mengubah teks menjadi format yang dapat diproses oleh model [19]. Proses pelatihan dilaksanakan selama 4 epoch dengan ukuran batch sebanyak 16.

Pemilihan algoritma ini didasarkan pada studi literatur sebelumnya. Algoritma *Naive Bayes* telah banyak digunakan dalam klasifikasi sentimen E-commerce karena kesederhanaannya dan kemampuannya dalam menangani teks pendek. Penelitian sebelumnya oleh [20] membuktikan bahwa *Naive Bayes* efektif untuk klasifikasi review aplikasi *Shopee* dengan akurasi mencapai 88%, namun mengalami kesulitan dalam membedakan sentimen netral.

Studi lain, oleh [21] berhasil menggunakan *Naive Bayes* dalam mengklasifikasikan ulasan produk Xiaomi di *Shopee* dengan akurasi 85%. Sementara, oleh [22] memperoleh akurasi sebesar 89% pada data Harbolnas *Shopee* menggunakan pendekatan serupa. Disisi lain, model Transformer seperti BERT mulai digunakan untuk mengatasi keterbatasan konteks.

Dalam penelitian oleh, [23] model Large Language Model (LLM) berbasis BERT digunakan untuk menganalisis 50.000 ulasan kategori Skincare di *Shopee* dengan akurasi mencapai 85,71%. Hal ini menunjukkan bahwa bert memiliki potensi yang sangat besar dalam menangani konteks kalimat yang kompleks serta variasi gaya Bahasa yang umum dan dapat ditemukan pada ulasan pengguna *Shopee*. Penelitian oleh [24] menunjukkan bahwa IndoBERT mampu mengungguli LSTM dengan F1-score 90.1% dan akurasi 90.6% pada klasifikasi ulasan e-commerce di Indonesia. Berdasarkan temuan tersebut, penelitian ini mengadopsi prosedur eksperimen yang serupa dengan menitikberatkan pada penggunaan model berbasis Transformer (BERT) untuk dibandingkan dengan metode klasifikasi tradisional yaitu *Naive Bayes*. Dengan demikian, pemilihan kedua metode ini didasarkan pada pertimbangan hasil penelitian sebelumnya sekaligus untuk menguji konsistensi performa pada dataset ulasan *Shopee* yang digunakan dalam penelitian ini.

Studi lain, oleh [25] menunjukkan bahwa integrasi metode BERT dengan Teknik pemrosesan berbasis atribut sehingga bisa menghasilkan peningkatan F1-Score sangat konsisten pada berbagai kategori produk. Oleh karena itu, pendekatan dalam penelitian ini tidak hanya menilai performa model dari segi akurasi umum, tetapi juga dapat menganalisis efektivitas klasifikasi pada ulasan yang dapat dipengaruhi oleh berbagai atribut produk seperti warna dan kategori.

2.5 Evaluasi Metriks

Dalam penelitian ini, model klasifikasi sentimen dibangun menggunakan algoritma *Naive Bayes* dan BERT untuk mengelompokkan ulasan pengguna *Shopee* ke dalam kelas sentimen seperti positif, negatif, berdasarkan teks ulasan, rating bintang, dan atribut produk seperti warna dan kategori. Evaluasi performa model dilakukan untuk mengetahui seberapa baik model dalam melakukan klasifikasi terhadap data. Beberapa metrik yang digunakan dalam penelitian ini meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Setiap metrik memiliki kegunaan masing-masing dalam mengevaluasi kinerja model secara menyeluruh [19].

a) Accuracy

Accuracy mengukur seberapa banyak prediksi model yang benar dibandingkan dengan seluruh jumlah data. Rumusnya adalah:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad [6]$$

b) Precision



Precision digunakan untuk mengukur seberapa akurat prediksi positif yang dilakukan oleh model. Metrik ini sangat berguna ketika konsekuensi dari prediksi positif yang salah (false positive) cukup besar. Rumus *precision* dapat dituliskan sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad [6]$$

c) *Recall*

Recall mengukur kemampuan model dalam mengenali seluruh data yang seharusnya diklasifikasikan sebagai positif. Metrik ini sangat penting pada kasus di mana kegagalan dalam mendeteksi kelas positif (false negative) dapat menimbulkan dampak besar. Rumusnya adalah:

$$Recall = \frac{TP}{TP+FN} \quad [6]$$

d) *F1-Score*

F1-score merupakan rata-rata harmonis dari *precision* dan *recall*. Metrik ini digunakan ketika diperlukan keseimbangan antara kedua metrik tersebut, terutama ketika distribusi kelas tidak seimbang. Rumus *F1-score* adalah sebagai berikut:

$$F1 - score = \frac{2*TP}{2*TP+FP+FN} \quad [6]$$

1. Visualisasi

Visualisasi dilakukan untuk menampilkan distribusi sentimen, korelasi antara kategori produk dan warna terhadap sentimen, serta performa model klasifikasi dalam bentuk confusion matrix.

2. Tools dan Lingkungan

Seluruh proses dilakukan dalam Jupyter Notebook menggunakan bahasa Python dengan library pendukung:

- a. Pandas, scikit-learn, matplotlib, seaborn
- b. Sastrawi, nltk (untuk NLP)
- c. transformers, datasets, torch (untuk BERT)

Lingkungan eksekusi: *Google Colab*.

3. HASIL DAN PEMBAHASAN

Penelitian ini mempunyai tujuan menganalisis sentimen ulasan produk pada platform *Shopee* dengan mengkaji hubungan antara rating yang diberikan pengguna dan atribut produk seperti warna atau kategori, serta membandingkan kinerja dari kedua model *Naive Bayes* dan BERT.

3.1 Pengumpulan Data

- a. Target Ulasan: 500 - 600 Data
- b. Total Ulasan: 2.300 Data

3.2 Preprocessing Data

- a. Cek Struktur & Missing Value

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2300 entries, 0 to 2299
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Customer ID           2300 non-null   int64
1   Ulasan Produk         2300 non-null   object
2   Rating                2300 non-null   int64
3   Warna                 2300 non-null   object
4   Varian / ukuran       2300 non-null   object
5   Kategori              2300 non-null   object
dtypes: int64(2), object(4)
memory usage: 107.9+ KB
None
Customer ID           0
Ulasan Produk         0
Rating                0
Warna                 0
Varian / ukuran       0
Kategori              0
dtype: int64
```

Gambar 4. Cek struktur dan missing value

b. Memberikan label sentimen berdasarkan rating.

Ulasan yang memperoleh rating bintang 4 dan 5 diberi label 1 yang berarti positif, sedangkan ulasan yang memiliki rating 2 dan 1 diberikan label 0 yang berarti negatif.

c. Case Folding

Merubah teks yang memiliki huruf kapital menjadi huruf kecil semua

3.3 Penerapan Model

Model

a. Akurasi Model: 90.47%

Akurasi diatas menunjukkan jumlah prediksi yang benar yaitu 570 data (TN+TP) dari total 630 data (baik dari kelas 1 maupun 0) dan terdapat kesalahan pada 58 data yang dimana 26 data kelas 0 diprediksi sebagai kelas 1 dan 34 data kelas 1 diprediksi sebagai kelas 0. Oleh karena itu, Model *Naive Bayes* masih cukup konsisten dalam memisahkan dua kelas meskipun ada kesalahan.

b. Classification Report:

1. Sentimen Negatif (0): Precision 0.88, Recall 0.91, F1-Score 0.90

a. Angka Precision untuk kelas 0 menunjukkan 0,88 dikarenakan Model memprediksi ada 292 review sebagai negatif (258 benar-benar negatif, 34 review seharusnya positif tapi diprediksi menjadi negative)

b. Angka Recall untuk kelas 0 menunjukkan 0,91 dikarenakan Ada 284 data yang benar-benar kelas 0. Model berhasil menangkap 258 di antaranya, hanya ada kesalahan yang dimana 26 adalah FN tetapi keliru dinyatakan positif (FP).

c. Angka F1-Score untuk kelas 0 menunjukkan 0,90 dikarenakan precision (0,88) lebih rendah dibandingkan Recall (0,91), maka F1-score seimbang diantara keduanya yaitu 0,90.

2. Sentimen Positif (1): Precision 0.92, Recall 0.90, F1-Score 0.91

a. Angka Precision untuk kelas 1 menunjukkan 0,92 dikarenakan Total prediksi "positif" = 312 (benar) + 26 (salah) = 338. Dari 338 prediksi positif, 312 benar (actual 1), dan 26 salah (actual 0 → FP). Sehingga Precision tinggi karena jumlah FP hanya 26, relatif kecil dibanding benar (312).

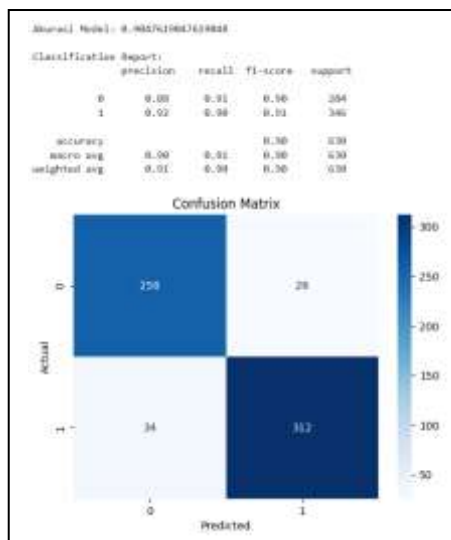
b. Angka Recall untuk kelas 1 menunjukkan 0,90 dikarenakan Total review yang benar-benar positif adalah 346 yang Dimana yang berhasil prediksi benar adalah 312, dan sisa nya 34 review positif salah diklasifikasikan sebagai negative (FN) sehingga Recall sedikit turun karena jumlah FN lebih besar daripada FP.

c. Angka F1-Score untuk kelas 1 menunjukkan 0,91 dikarenakan Precision tinggi sebesar 0,92 dan Recall sedikit lebih rendah 0,90 sehingga F1-Score berada di tengah yaitu 0,91.

Confusion Matrix:

- True Negative (0 diklasifikasikan sebagai 0):258
- False Positive (0 diklasifikasikan sebagai 1):26
- False Negative (1 diklasifikasikan sebagai 0): 34
- True Positive (1 diklasifikasikan sebagai 1): 312

Model menunjukkan performa cukup baik dengan F1-Score mendekati 0.90 untuk kedua kelas. Namun, jumlah prediksi yang salah (FP dan FN) masih cukup signifikan, terutama pada sentimen positif yang salah diklasifikasikan sebagai negatif.



Gambar 5. Confusion Matrix dan Hasil Evaluasi *Naive Bayes*

Dari gambar 5 memperlihatkan hasil evaluasi model *Naive Bayes* dalam klasifikasi sentimen. Dari confusion matrix terlihat bahwa model menghasilkan 312 True Positive (TP) dan 258 True Negative (TN), sementara masih terdapat 26 False Positive (FP) dan 34 False Negative (FN). Nilai akurasi keseluruhan mencapai 0,90 dengan precision sebesar 0,88 untuk kelas negatif dan 0,92 untuk kelas positif.

	Predicted Positive	Predicted Negative
Actual Positive (1)	TP= 312	FN=34
Actual Negative (0)	FP=26	TN=258

Dimana:

- Class 1 = Positive (Review Positif)
- Class 0 = Negative (Review Negatif)

Kinerja Confusion Matrix

- True Positive (TP) menunjukkan bahwa review adalah positif dan prediksi juga menyatakan bahwa review tersebut positif.
- True Negative (TN) menunjukkan bahwa review adalah negatif dan prediksi juga menyatakan bahwa review tersebut negatif.
- False Positive (FP) menunjukkan bahwa review adalah negatif, tetapi prediksinya menyatakan bahwa review tersebut positif.
- False Negative (FN) menunjukkan bahwa review adalah positif, namun prediksinya menyatakan bahwa review tersebut negatif.

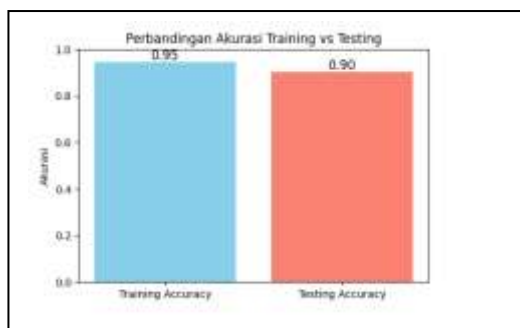
Grafik Akurasi Model *Naive Bayes*

Akurasi Training tinggi: 0.95 (95%)

Akurasi Testing: 0.90 (90%)

Dari kedua akurasi diatas terlihat ada selisih sebesar 5% antara akurasi Training dan testing. Ini menunjukkan model berfungsi dengan baik pada data pelatihan, tetapi tetap mempertahankan performa yang baik pada data yang belum pernah dilihat (testing)

Dari kedua akurasi diatas terdapat indikasi overfitting yaitu selisih 5% antara akurasi Training dan testing tidak terlalu besar, sehingga belum menunjukkan overfitting yang parah dan model masih mampu melakukan generalisasi terhadap data baru dengan cukup baik.



Gambar 6. Grafik Akurasi Model Naive Bayes

Model BERT

1. Akurasi Model sebesar 94.92%

Dari total data 630 (TN+TP+FN+FP) 598 diprediksi dengan benar (273 dari kelas 0 dan 325 dari kelas 1) terdapat 32 data yang salah diprediksi diantaranya 11 data kelas 0 salah diklasifikasikan sebagai 1, 21 data kelas 1 salah diklasifikasikan sebagai 0, sehingga BERT sangat akurat dalam klasifikasi karena terdapat proporsi kesalahan data sangat kecil (32 dari 630 data). Oleh karena itu, BERT mampu mengenali hampir seluruh data dengan benar.

2. Classification Report:

1. Sentimen Negatif (0): Precision 0.93, Recall 0.96, F1-Score 0.94

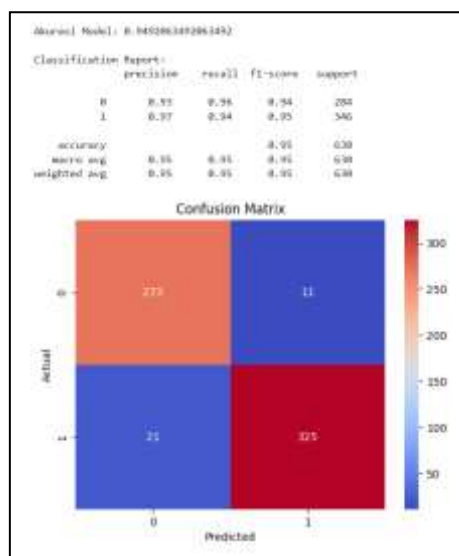
- Angka Precision untuk kelas 0 menunjukkan 0,93 dikarenakan Dari semua data yang diprediksi sebagai kelas 0, sebanyak 93% memang benar-benar kelas 0. Ada sedikit "False Positive" (kelas 1 yang salah dikira kelas 0).
- Angka Recall untuk kelas 0 menunjukkan 0,96 dikarenakan Dari semua data yang sebenarnya kelas 0, model berhasil mengenali 96% dengan benar. Hanya 11 salah dikira sebagai kelas 1.
- Angka F1-Score untuk kelas 0 menunjukkan 0,94 dikarenakan F1 mencari keseimbangan antara precision dan recall dan disini dari precision dan Recall sama-sama tinggi dan mendekati satu sama lain, maka F1 juga tinggi.

2. Sentimen Positif (1): Precision 0.97, Recall 0.94, F1-Score 0.95

- Angka Precision untuk kelas 1 menunjukkan 0,97 dikarenakan Model mampu memprediksi 336 data sebagai kelas 1, dari 336 diantaranya 325 sebagai True Postif, dan 11 sebagai false positif sehingga precision sangat tinggi sebagai 0,97.
- Angka Recall untuk kelas 1 menunjukkan 0,94 dikarenakan Total pada data kelas 1 adalah 346, dari 346 diantaranya 325 berhasil mengenali model (TP), dan 21 sebagai gagal dikira kelas 0 sehingga hasil recall masih tinggi.
- Angka F1-Score untuk kelas 1 menunjukkan 0,95 dikarenakan Precision sangat tinggi sebesar 0,97 dan recall sedikit lebih rendah 0,94 sehingga F1-Score berada diantara kedua yaitu 0,95.

Confusion Matrix:

Bert memberikan peningkatan akurasi dan ketepatan klasifikasi secara keseluruhan dibandingkan . F1-Score yang lebih tinggi menandakan model lebih seimbang dalam mendeteksi kedua kelas.



Gambar 7. Confusion Matrix dan Hasil Evaluasi BERT

Dari Gambar 7 memperlihatkan hasil evaluasi model BERT dalam klasifikasi sentiment. Dari confusion matrix terlihat bahwa model menghasilkan 325 True Positive (TP) dan 273 True Negative (TN), sementara masih terdapat 11 False Positive dan 21 False Negative (FN). Nilai akurasi keseluruhan mencapai 0,94 dengan precision sebesar 0,93 untuk kelas Negatif dan 0,97 untuk kelas Positif.

	Predicted Positive	Predicted Negative
Actual Positive (1)	TP= 325	FN=21
Actual Negative (0)	FP=11	TN=273

Dimana:

Class 1 = Positive (Review Positif)

Class 0 = Negative (Review Negatif)

Kinerja Confusion Matrix

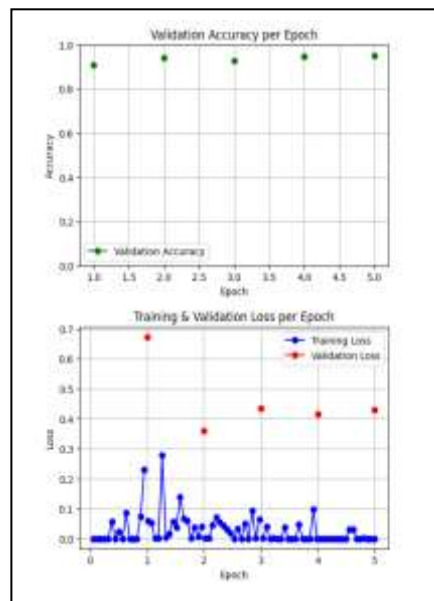
- True Positive* (TP) menunjukkan bahwa review positif dan prediksi menyatakan positif. Artinya, sistem berhasil mengenali review yang memang bernada positif.
- True Negative* (TN) menunjukkan bahwa review negatif dan prediksi menyatakan negatif. Artinya, sistem mampu mengenali review negatif dengan tepat.
- False Positive* (FP) menunjukkan review negatif, tetapi prediksinya menyatakan positif. Artinya, sistem salah menilai ulasan negatif sebagai positif.
- False Negative* (FN) menunjukkan review positif, namun prediksinya menyatakan negatif. Artinya, sistem gagal mengenali ulasan yang sebenarnya bernada positif.

Grafik Akurasi dan Loss pada Bert

Grafik Validation Accuracy menunjukkan bahwa akurasi validasi terlalu tinggi di setiap epoch yaitu diatas 90% bahkan mendekati 95-97% dan ini menunjukkan bahwa model Bert mampu menjaga performa yang baik terhadap data yang belum pernah dilihat (data validasi)

Grafik Loss

Training loss tampak sangat rendah dan fluktuatif mendekati nol. Validation loss terlihat cukup tinggi (sekitar 0.35-0.68). Model Bert tidak mengalami underfitting karena Accuracy sudah tinggi, namun ada terindikasi awal overfitting ringan, yang ditandai dengan Training loss yang sangat rendah dan Validation Loss cukup tinggi.



Gambar 8. Hasil Evaluasi BERT

Analisis Sentimen berdasarkan kategori dan warna produk

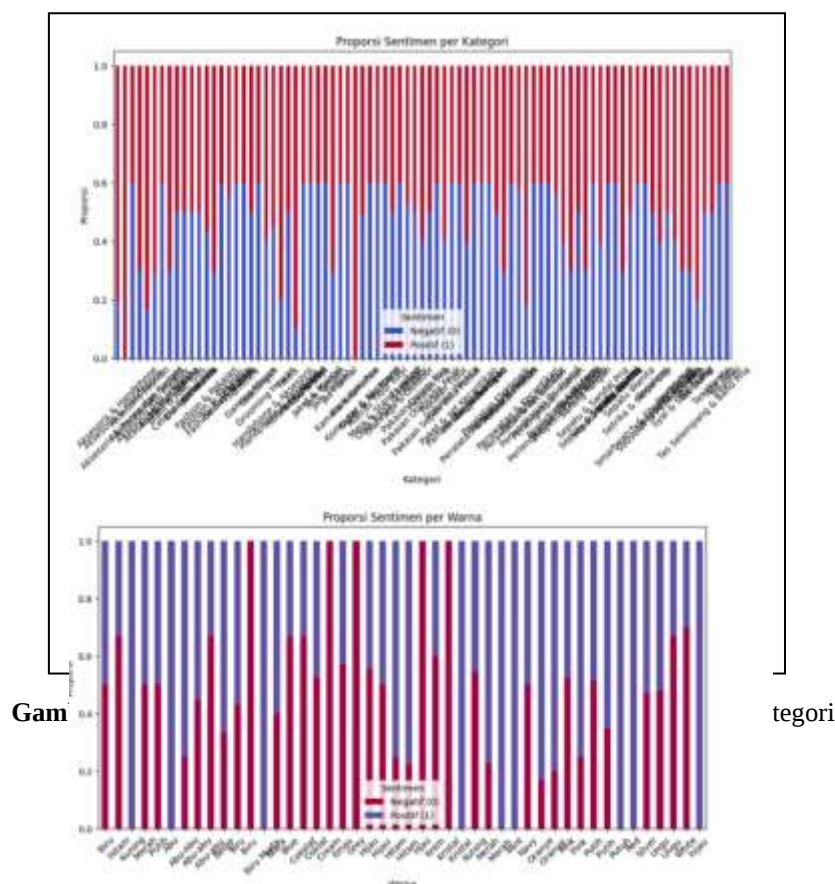
a.

1. Kategori Produk:

Sebagian besar kategori menunjukkan proporsi sentimen positif yang tinggi. Namun terdapat fluktuasi besar antara kategori; beberapa kategori memiliki distribusi sentimen negative yang dominan.

2. Warna Produk:

Warna seperti “Abu-abu”, “Hitam”, dan “Putih” cenderung menghasilkan sentimen negative lebih tinggi dan warna seperti “Merah”, “Pink”, dan “Biru” didominasi oleh sentimen positif.



Gam

tegori

b. Bert**1. Kategori Produk:**

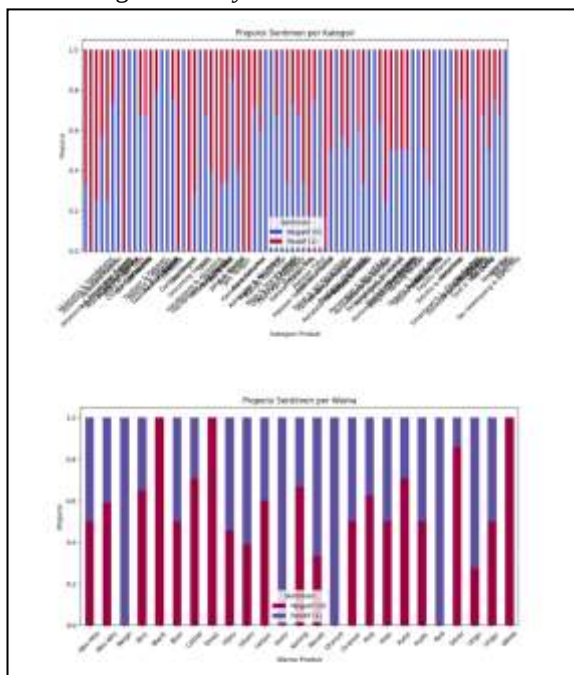
Bert mengklasifikasikan ulasan dengan distribusi yang lebih stabil dan proporsi sentimen positif yang lebih dominan hampir di semua kategori.

Kategori seperti “Perlengkapan Mandi” dan “Tas Slempong” menunjukkan distribusi sentimen paling seimbang.

2. Warna Produk:

Model Bert lebih konsisten dalam mengidentifikasi sentimen positif pada sebagian besar warna.

Warna seperti “Hitam”, “Coklat”, dan “Hijau” tetap menunjukkan kecenderungan lebih tinggi terhadap sentimen negatif, tetapi lebih terkontrol dibanding *Naive Bayes*.



Gambar 10. Analisis BERT berdasarkan atribut produk Warna/kategori

4. KESIMPULAN

Berdasarkan hasil evaluasi yang telah dilakukan, model BERT terbukti lebih efektif dibandingkan *Naive Bayes* dalam melakukan klasifikasi sentimen pada ulasan produk di *Shopee*. Keunggulan ini terlihat dari nilai akurasi yang lebih tinggi, yaitu 94,9% untuk BERT dibandingkan dengan 90,4% pada *Naive Bayes*. Selain itu, BERT juga memiliki nilai precision, recall, dan F1-score yang lebih baik, menandakan performa klasifikasi yang lebih seimbang dan akurat. Dari analisis confusion matrix, BERT menghasilkan jumlah kesalahan klasifikasi yang lebih sedikit, baik dalam hal false positives maupun false negatives, dibandingkan dengan *Naive Bayes*. Implikasi praktis dari temuan ini menunjukkan bahwa BERT berpotensi untuk diimplementasikan dalam sistem rekomendasi *Shopee*, misalnya dengan memberikan rekomendasi produk berdasarkan ulasan positif yang lebih akurat atau menyaring produk dengan ulasan negatif yang konsisten. Selain itu, keunggulan BERT tidak hanya relevan untuk *Shopee*, tetapi juga dapat diterapkan pada platform e-commerce lain yang memiliki karakteristik ulasan serupa. Dengan demikian, penelitian ini memberikan kontribusi praktis bagi pengembangan sistem analitik berbasis sentimen dalam mendukung pengambilan keputusan strategis di berbagai layanan e-commerce. Analisis lebih lanjut terhadap atribut produk seperti kategori dan warna menunjukkan bahwa model *Naive Bayes* cenderung menghasilkan distribusi sentimen yang tidak stabil, sedangkan BERT mampu menangkap konteks ulasan yang lebih rumit sehingga memberikan hasil klasifikasi yang lebih konsisten dan proporsional. Oleh karena itu, model BERT lebih direkomendasikan untuk digunakan dalam analisis sentimen ulasan produk *Shopee*, khususnya karena kemampuannya untuk memahami konteks kalimat dengan lebih mendalam. Namun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama adalah data yang digunakan hanya terbatas pada ulasan produk dari satu platform e-commerce (*Shopee*) sehingga generalisasi ke platform lain masih perlu diuji. Kedua adalah atribut produk yang dianalisis masih terbatas pada kategori dan warna, sementara faktor lain seperti harga, kualitas pengiriman, dan lokasi penjual belum dipertimbangkan. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas cakupan data ke berbagai platform e-commerce yang lain serta menambahkan atribut produk lain agar hasil analisis menjadi lebih komprehensif. Dataset juga dapat diperluas untuk meningkatkan akurasi dan generalisasi model. Selain *Naive Bayes* dan BERT, algoritma lain seperti LSTM, Random Forest, atau DistilBERT



dapat diuji untuk perbandingan performa yang lebih luas. Dengan demikian, hasil penelitian ini tidak hanya mendukung penerapan BERT untuk analisis sentimen di platform e-commerce seperti *Shopee*, tetapi juga dapat menjadi acuan bagi pengembangan sistem serupa di e-commerce lain di Indonesia. Penanganan data multibahasa dan pendekatan multi-label classification juga perlu dipertimbangkan, terutama pada ulasan yang kompleks. Selain itu, hasil analisis sentimen dapat diintegrasikan ke dalam sistem rekomendasi produk guna meningkatkan pengalaman pengguna di platform e-commerce.

REFERENCES

- [1] Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, and Anis Fitri Nur Masruriyah, "Analisis sentimen ulasan pada e-commerce *Shopee* menggunakan algoritma *Naive Bayes* dan support vector machine," *INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 16–26, 2023, doi: 10.37373/infotech.v4i1.422.
- [2] S. Aras, R. Ruimassa, E. Agustinus, B. Wambrau, and E. B. Palalangan, "Sentiment Analysis on *Shopee* Product Reviews Using IndoBERT," vol. 6, no. 3, pp. 1616–1627, 2024, doi: 10.51519/journalisi.v6i3.814.
- [3] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma *Naive Bayes* untuk Analisis Sentimen Ulasan *Shopee* pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [4] Y. Romadhoni and K. F. H. Holle, "Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode *Naive Bayes* dan LSTM," *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 118–124, 2022, doi: 10.30591/jpit.v7i2.3191.
- [5] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode *Naive Bayes* Dan Knn," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020, doi: 10.33480/inti.v15i1.1347.
- [6] S. G. Alexander, A. T. Ananto, I. P. Adhitya, P. Mangku, and B. Liano, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Konten Deepfake Tokoh Publik," *KAKIFIKOM (Kumpulan Artik. Karya Ilm. Fak. Ilmu Komputer)*, vol. 05, no. 02, pp. 95–102, 2023.
- [7] A. M. Putri, W. Khafa Nofa, D. Anggraini, and P. Hapsari, "Penerapan metode bert untuk analisis sentimen ulasan pengguna aplikasi segari di google play store," vol. 4, no. 1, pp. 89–104, 2025.
- [8] A. Ernawati, A. O. Sari, S. N. Sofyan, M. Iqbal, and R. F. W. Wijaya, "Implementasi Algoritma *Naive Bayes* dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan," *Bull. Inf. Technol.*, vol. 4, no. 4, pp. 533–543, 2023, doi: 10.47065/bit.v4i4.1090.
- [9] M. Xanderina *et al.*, "Analisis Sentimen Ulasan E-Commerce *Shopee* Pada Google Play Store Menggunakan Machine Learning," *J. Eng. Sustain. Technol.*, vol. 10, no. 2, pp. 990–998, 2024, [Online]. Available: <https://ejournal.unma.ac.id/index.php/j-ensitec/article/view/9071>
- [10] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [11] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," *J. Manaj. Inform.*, vol. 12, no. 2, pp. 89–99, 2022, doi: 10.34010/jamika.v12i2.7764.
- [12] D. R. A. and Y. F. Riti, "Analisis Sentimen Ulasan Aplikasi Genshin Impact di Play Store Menggunakan Random Forest," pp. 578–586, 2024.
- [13] ryan putranda kristianto, Hendra, and S. S. Tandjung, "Studi Komparatif: Performansi Akurasi Algoritma Klasifikasi untuk Analisis Sentimen Pada Kandidat Presiden RI di Pemilu 2024," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 13, no. 4, Oct. 2024, doi: 10.30591/smartcomp.v13i4.6286.
- [14] M. Iqbal, A. D. Wiranata, R. Suwito, R. F. Ananda, U. Muhammadiyah, and P. Hamka, "Perbandingan Algoritma *Naive Bayes*, KNN, dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter," vol. 4, no. 3, pp. 1799–1807, 2023, doi: 10.30865/klik.v4i3.1402.
- [15] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi *Shopee* Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [16] A. Oktian Permana and Sudin Saepudin, "Perbandingan algoritma k-nearest neighbor dan *Naive Bayes* pada aplikasi *Shopee*," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 1, pp. 25–32, 2023, doi: 10.37859/coscitech.v4i1.4474.
- [17] A. H. Hasugian, M. Fakhri, and D. Zukhoiriyah, "Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma *Naive Bayes*," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 98, 2023, doi: 10.53513/jsk.v6i1.7400.
- [18] T. Fardaningsih, A. Lutfiyani, T. Informatika, and J. Tengan, "Perbandingan Analisis Sentimen Pengguna Aplikasi *Shopee* dan Lazada pada Situs Google Play Store Menggunakan Algoritma K-Nearest Neighbor dan *Naive Bayes*," vol. 4, no. 3, pp. 563–578, 2025, doi: 10.55123/insologi.v4i3.5646.
- [19] E. Subowo and I. K. Negeri Saifuddin Zuhri Purwokerto edysubowo, "Implementasi Pembelajaran Mendalam



- dalam Klasifikasi Sentimen Ulasan Aplikasi: Evaluasi Model BERT, LSTM, dan CNN,” 2024.
- [20] R. R. Rismansyah, A. Sudiarjo, and T. Mufizar, “Analisis Sentimen Ulasan *Shopee* Pada Google Play Store Menggunakan Algoritma *Naive Bayes*,” *J. Elektro Inform. Swadharma*, vol. 5, no. 1, pp. 109–119, 2025.
- [21] L. O. Sihombing, H. Hannie, and B. A. Dermawan, “Sentimen Analisis Customer Review Produk *Shopee* Indonesia Menggunakan Algoritma *Naive Bayes Classifier*,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.
- [22] L. M. Siniwi, A. Prahutama, and A. R. Hakim, “QUERY EXPANSION RANKING PADA ANALISIS SENTIMEN MENGGUNAKAN KLASIFIKASI MULTINOMIAL *NAÏVE BAYES* (Studi Kasus : Ulasan Aplikasi *Shopee* pada Hari Belanja Online Nasional 2020),” *J. Gaussian*, vol. 10, no. 3, pp. 377–387, 2021, doi: 10.14710/j.gauss.v10i3.32795.
- [23] R. A. Yunita and E. Indra, “Analisis Sentimen Ulasan Produk Menggunakan Large Language Models : Studi Kasus Pada *Shopee*,” *J. Sist. Inf. dan Ilmu Komput. Prima*, vol. 8, no. 2, pp. 1–13, 2025.
- [24] A. Kamal and R. Astri, “Eksplorasi Sentimen Pengguna pada Aplikasi E-Commerce dengan Deep Learning,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 7, no. 3, pp. 435–441, Jul. 2025, doi: 10.47233/jteksis.v7i3.2010.
- [25] M. Saifurridho, M. Martanto, and U. Hayati, “Analisis Algoritma K-Nearest Neighbor terhadap Sentimen Pengguna Aplikasi *Shopee*,” *J. Inform. Terpadu*, vol. 10, no. 1, pp. 21–26, 2024, doi: 10.54914/jit.v10i1.1054.