

PERBANDINGAN KINERJA NAÏVE BAYES DAN INDOBERT DALAM KLASIFIKASI SENTIMEN ULASAN PELANGGAN PADA APLIKASI SHOPEE

Arridho Ramadhan Firdaus¹, Deyana Kusuma Wardani², Suhendra³

^{1,3} Teknologi Rekayasa Perangkat Lunak, Politeknik Astra

² Manajemen Informatika, Politeknik Astra

Jl. Gaharu Blok F3 Delta Silicon II Cibatu, Cikarang Selatan Kabupaten Bekasi Jawa Barat, Indonesia

arridho.firdaus@polytechnic.astra.ac.id

ABSTRAK

Perkembangan teknologi digital mendorong peningkatan penggunaan *e-commerce*, seperti Shopee, yang memiliki banyaknya jumlah ulasan pelanggan. Kondisi tersebut menjadikan analisis sentimen sebagai salah satu pendekatan komputasional yang dapat digunakan untuk mengidentifikasi polaritas sentimen dan memperoleh kesimpulan opini dari pelanggan terhadap layanan *e-commerce*. Namun, karena banyaknya ambiguitas dalam teks berbahasa Indonesia dan lonjakan jumlah ulasan pelanggan menjadi tantangan utama dalam menentukan algoritma klasifikasi sentimen yang mampu memahami ulasan secara efektif. Naïve Bayes merupakan salah satu algoritma klasifikasi yang banyak digunakan karena memiliki komputasi yang relatif ringan, sedangkan IndoBERT memiliki keunggulan dalam memanfaatkan konteks dan hubungan antar kata dalam kalimat. Dari kedua algoritma tersebut, penelitian ini menguji dan membandingkan kinerja dari Naïve Bayes dan IndoBERT untuk melakukan klasifikasi polaritas sentimen positif dan negatif pada ulasan pengguna aplikasi *e-commerce*. Penelitian ini menggunakan dataset ulasan pengguna berbahasa Indonesia pada Shopee. Hasil pengujian menunjukkan bahwa IndoBERT memperoleh akurasi sebesar 86.5%, sedangkan Naïve Bayes memperoleh akurasi sebesar 80%. Dari hasil tersebut, IndoBERT memiliki kinerja klasifikasi yang lebih tinggi dibandingkan Naïve Bayes dalam mengidentifikasi polaritas sentimen serta mampu memahami hubungan antarkata yang dapat mendukung peningkatan kinerja analisis sentimen pada ulasan berbahasa Indonesia.

Kata kunci : Analisis Sentimen, E-Commerce, Natural Language Processing, Naïve Bayes, IndoBERT

1. PENDAHULUAN

Perkembangan teknologi digital yang pesat, terutama sejak pandemi COVID-19 pada tahun 2019, telah mengubah pola masyarakat dalam melakukan aktivitas belanja [1]. Berbagai platform *e-commerce* semakin banyak digunakan oleh masyarakat Indonesia, salah satunya Shopee yang memiliki jumlah pengguna dan ulasan yang banyak. Melalui aplikasi tersebut, pengguna dapat memberikan ulasan, penilaian, maupun keluhan terkait produk yang dijual maupun layanan dengan mudah. Ulasan pengguna umumnya sebagai cerminan luapan emosi secara spontan dengan penggunaan bahasa sehari-hari atau tidak baku. Berdasarkan informasi dari Google Play Store, aplikasi Shopee telah memperoleh lebih dari 18 juta ulasan, sehingga jumlah data yang besar tersebut tidak memungkinkan untuk dianalisis secara manual. Oleh karena itu, diperlukan pendekatan berupa analisis sentimen yang terdapat dalam ulasan pengguna Shopee.

Cabang dari ilmu kecerdasan buatan yang mempelajari bahasa manusia adalah *Natural Language Processing* (NLP) [2]. Salah satu penerapan NLP yang sering digunakan adalah analisis sentimen, yaitu metode untuk mengidentifikasi maupun mengelompokkan opini atau emosi pengguna ke dalam kategori polaritas, seperti positif, negatif, dan netral [3]. Metode ini dapat membantu perusahaan atau pemilik bisnis untuk memahami tingkat kepuasan pelanggan secara efisien tanpa harus membaca seluruh ulasan yang ada.

Ketika mengimplementasikan analisis sentimen pada ulasan teks, pemilihan algoritma klasifikasi menjadi hal yang krusial untuk mempelajari pola data teks yang kompleks. Algoritma *Naïve Bayes* yang dipadukan dengan metode pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) menjadi salah satu algoritma yang sering digunakan. Algoritma ini menghitung probabilitas suatu dokumen ulasan untuk dapat masuk kedalam kategori tertentu berdasarkan frekuensi kemunculan kata-kata di dalamnya [4]. Keunggulan dari *Naïve Bayes* adalah proses komputasi yang tergolong ringan serta waktu pelatihan model yang relatif cepat [5]. Namun, algoritma ini tidak memperhatikan hubungan antara kata dalam sebuah kalimat, sehingga sering kali memicu kesalahan klasifikasi pada teks ulasan, terutama pada kalimat yang terdapat kata ambigu atau struktur kata yang tidak baku.

Algoritma IndoBERT menjadi salah satu algoritma yang saat ini sedang populer untuk menganalisa sentimen dari ulasan. Berbeda dengan Naïve Bayes, IndoBERT tidak hanya membaca teks dari satu arah, melainkan menganalisa makna dari sebuah kata dengan mempertimbangkan konteks dari semua kata yang ada dalam satu kalimat [6]. Hal ini menjadikan algoritma IndoBERT tidak hanya fokus pada kemunculan kata, tetapi juga mempertahankan informasi semantik dari keseluruhan kalimat.

Berdasarkan dari perbedaan kedua algoritma yang signifikan, penelitian ini bertujuan untuk menguji dan membandingkan kinerja dari Naïve Bayes dan

IndoBERT dalam melakukan klasifikasi sentimen positif dan negatif, dengan mengecualikan sentimen netral. Hasil analisis perbandingan ini diharapkan memberikan wawasan mengenai seberapa besar performa dari kedua algoritma tersebut dalam menangani data teks ulasan *e-commerce*.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu menunjukkan bahwa pemodelan *Natural Language Processing* (NLP) menggunakan arsitektur Transformer, seperti IndoBERT, memiliki keunggulan dalam menangkap konteks semantik teks berbahasa Indonesia. Beberapa peneliti telah membuktikan keunggulan dari IndoBERT yang menghasilkan capaian akurasi yang tinggi, seperti pada penelitian tentang klasifikasi opini publik terhadap kebijakan pendidikan yang dapat mencapai akurasi hingga 98% [7]. Kemampuan IndoBERT juga dapat diimplementasikan untuk analisis sentimen dari berbagai aspek pada aplikasi layanan publik digital, dengan akurasi diatas 97% [8]. Selain itu, performa IndoBERT dapat dioptimalkan melalui penggabungan ekstraksi fitur GloVe yang dapat mencapai akurasi 91% pada data dari ulasan aplikasi *zoom* [9]. Dari ketiga penelitian ini menjelaskan bahwa IndoBERT merupakan model yang sangat baik dalam mengenali pola bahasa yang rumit, termasuk kata tidak baku dan bahasa gaul.

Di sisi lain, algoritma klasifikasi probabilitas, seperti Naïve Bayes juga memberikan performa yang sangat andal dalam melakukan klasifikasi teks, terutama ketika digabungkan dengan teknik pembobotan kata *Term Frequency-Inverse Document* (TF-IDF). Berbagai penelitian membuktikan bahwa Naïve Bayes dapat mencapai akurasi yang maksimal dalam klasifikasi teks, diantaranya capaian akurasi sebesar 90% dalam mengevaluasi opini masyarakat terhadap layanan rumah sakit [10], serta akurasi 81% dalam menganalisis ulasan pada aplikasi penggalangan dana sosial [11]. Implementasi Naïve Bayes pada ulasan dompet digital tidak hanya menghasilkan akurasi algoritma sebesar 88.5% saja, namun juga menghasilkan waktu komputasi prediksi 129 kali lebih cepat dibandingkan algoritma pembandingnya [12]. Hal ini membuktikan bahwa Naïve Bayes tetap menjadi metode klasifikasi yang efisien dan stabil untuk memproses dataset teks berskala besar. Dari penelitian sebelumnya, IndoBERT memiliki kemampuan dalam memahami konteks kalimat, sedangkan Naïve Bayes memiliki efisiensi komputasi yang baik. Oleh karena itu, penelitian ini akan melakukan perbandingan antara kedua algoritma tersebut pada data ulasan aplikasi *e-commerce*.

2.2. Natural Language Processing (NLP)

Natural Language Processing, cabang ilmu dari kecerdasan buatan yang memiliki peran penting sebagai penghubung antara bahasa manusia yang

cenderung ambigu dengan sistem komputer [13]. Fokus utamanya adalah algoritma yang dapat membaca, menguraikan, hingga memanipulasi bahasa manusia. Pada awal perkembangan, NLP menggunakan aturan tata bahasa (*rule-based*) tertentu [14]. Namun, seiring berjalannya waktu, pendekatan tersebut mulai tergantikan menjadi berbasis data dengan memanfaatkan *machine learning* dan *deep learning* agar sistem dapat memahami konteks yang ada dalam sebuah kalimat.

Data yang sering digunakan dalam penelitian di bidang ini adalah teks ulasan (*review*). Teks ulasan menjadi cerminan dari suara konsumen yang spontan dan emosional, serta sering kali menggunakan bahasa gaul (slang) dan bahasa yang tidak baku [15]. Oleh karena itu, kemampuan NLP dalam memahami konteks dapat membantu mengidentifikasi kata atau frasa yang kurang bermakna dalam sebuah ulasan. Proses tersebut juga membantu menganalisa ribuan ulasan secara efisien, sehingga informasi dan kesimpulan dapat diperoleh tanpa perlu membaca setiap ulasan secara manual [16].

Pengolahan teks ulasan menggunakan NLP dapat dilakukan melalui berbagai pendekatan, salah satunya adalah analisis sentimen. Analisis sentimen memanfaatkan kemampuan NLP untuk mengidentifikasi, mengekstraksi, dan mengelompokkan opini atau emosi yang terdapat dalam teks ke dalam ulasan positif, netral, atau negatif [17]. Tahap awal perkembangannya, analisis sentimen banyak menggunakan pendekatan *lexicon-based* dengan mencocokkan kata dalam teks dengan kamus sentimen [18]. Seiring perkembangan NLP, analisis tidak lagi terbatas pada pencocokan kata, tetapi juga dapat mempertimbangkan konteks dan topik yang dibahas dalam suatu ulasan. Sebagai contoh, ulasan mengenai produk kosmetik dapat dianalisis berdasarkan beberapa topik, seperti harga, bahan, dan efek pemakaian.

2.3. Tahapan Pra-Pemrosesan Teks (Preprocessing)

Pra-pemrosesan teks (*Text Preprocessing*) merupakan tahapan penting dalam *Natural Language Processing* yang bertujuan mengubah data teks mentah menjadi bentuk yang lebih terstruktur dan dapat diolah oleh mesin [19]. Pada dasarnya, mesin tidak dapat memahami teks secara langsung seperti manusia, sehingga teks perlu diubah ke dalam bentuk yang dapat diproses oleh sistem. Data mentah biasanya masih terdapat berbagai *noise*, seperti *typo*, singkatan atau bahasa gaul, penggunaan tanda baca berlebihan, dan bentuk format penulisan yang tidak konsisten [20]. Oleh karena itu, tahapan ini dilakukan untuk mengurangi *noise* tersebut, sehingga menghasilkan data yang lebih rapi dan siap digunakan untuk algoritma pemodelan.

Secara umum, pra-pemrosesan teks terdiri dari beberapa tahapan yang dilakukan secara berurutan, mulai dari menyeragamkan teks ke dalam huruf kecil, membersihkan data dengan menghilangkan karakter

yang tidak diperlukan, memecahkan teks menjadi beberapa unit leksikal, menghilangkan kata-kata yang kurang memiliki makna, hingga mengubah teks menjadi bentuk dasar dengan menghilangkan imbuhan [21]. Namun penerapan setiap tahapan dapat disesuaikan dengan kebutuhan dan karakteristik model yang digunakan.

2.4. Algoritma Klasifikasi Teks

Pada analisa sentimen, algoritma klasifikasi teks digunakan untuk mengelompokkan teks berdasarkan polaritas tertentu, seperti positif, negatif, atau netral [3]. Sebelum proses klasifikasi dilakukan, teks yang telah melalui tahap pembersihan data perlu diubah menjadi representasi numerik agar dapat diproses oleh komputer. Selanjutnya, algoritma klasifikasi mempelajari pola dari fitur-fitur numerik tersebut untuk mengenali karakteristik masing-masing kategori. Pola yang dipelajari kemudian digunakan untuk memprediksi label pada data teks baru yang belum pernah diproses sebelumnya.

Salah satu algoritma klasik yang sering digunakan dalam klasifikasi teks adalah Naive Bayes. Algoritma ini bekerja dengan menghitung probabilitas suatu dokumen untuk masuk ke dalam kategori tertentu berdasarkan kemunculan kata-kata di dalam dokumen tersebut [4]. Selain itu, algoritma ini juga dapat memberikan hasil yang cukup baik ketika dikombinasikan dengan metode pembobotan sederhana.

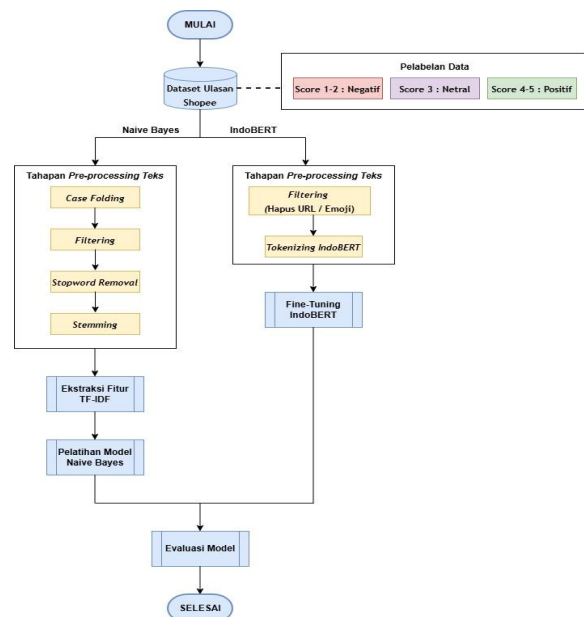
Pada dasarnya, Naive Bayes tidak mempertimbangkan hubungan antar kata dalam sebuah kalimat. Oleh karena itu, diperlukan pendekatan lain yang mampu menangkap hubungan dan konteks antar kata, salah satunya IndoBERT. Algoritma ini memiliki kemampuan memahami bahasa Indonesia melalui proses *pre-training* menggunakan korpus teks berbahasa Indonesia dalam jumlah besar [22]. Melalui pendekatan *bidirectional*, IndoBERT dapat memahami sebuah kata dengan mempertimbangkan konteks dari kata-kata yang berada di sisi kiri dan kanan secara bersamaan [23]. Hasilnya IndoBERT tidak hanya memperhatikan kemunculan kata, tetapi juga dapat memahami makna kata berdasarkan konteks kalimat secara keseluruhan.

2.5. Evaluasi Model

Evaluasi performa algoritma klasifikasi sentimen dapat menggunakan *confusion matrix* [24]. Pada *confusion matrix*, performa klasifikasi dipecah menjadi empat kuadran utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Area *True Positive* dan *True Negative* menunjukkan kondisi ketika model berhasil menebak sentimen secara akurat, misalnya ulasan positif sebagai positif dan ulasan negatif sebagai negatif. Sebaliknya, area *False Positive* dan *False Negative* mempresentasikan kesalahan algoritma yang digunakan [25], seperti ketika ulasan complain dianggap sebagai pujian oleh sistem.

3. METODE PENELITIAN

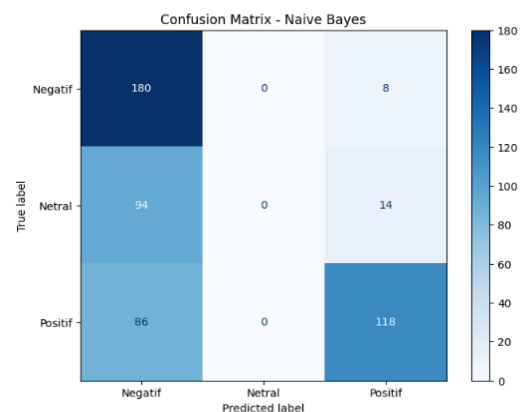
Metode penelitian dirancang secara sistematis yang digambarkan pada Gambar 1.



Gambar 1. Alur Diagram dari Penelitian

3.1. Pengumpulan Dataset

Data yang digunakan dalam penelitian ini diperoleh dari *dataset* publik yang berasal dari Kaggle, yaitu *E-Commerce Sampled Reviews in Bahasa Indonesia #1*. Atribut utama yang diambil meliputi konten dan skor. Ulasan dari *dataset* tersebut direpresentasikan dalam bentuk teks berbahasa Indonesia yang mencerminkan opini, tingkat kepuasan, serta pengalaman dari pengguna terhadap layanan aplikasi *e-commerce*. Nilai skor yang berskala 1 hingga 5, dikonversikan ke dalam tiga kategori label kelas sentimen, yaitu skor 1 dan 2 kategori sentimen negatif, kelas 3 kategori sentimen netral, dan kelas 4 dan 5 kategori kelas positif. Ulasan dengan skor 3 (netral) tidak disertakan dalam pemodelan lebih lanjut. Penelitian ini difokuskan pada klasifikasi positif dan negative dengan menghindari ambiguitas kontekstual yang sering kali muncul pada teks ulasan.



Gambar 2. Jumlah Sebaran Data dalam 3 (tiga) Kategori Sentimen

3.2. Pra-Pemrosesan Teks

Penelitian ini membandingkan dua pendekatan algoritma yang berbeda, Naïve Bayes dan IndoBERT, maka tahapan pra-pemrosesan disesuaikan dengan karakteristik masing-masing algoritma sebagai berikut:

3.3. Pra-Pemrosesan untuk Naïve Bayes

Algoritma Naïve Bayes menggunakan representasi dari fitur kata, sehingga memerlukan tahapan pembersihan yang lebih banyak. Tahapan ini meliputi:

- Case Folding,**
Tahapan ini merupakan awalan dari pra-pemrosesan teks dengan mengubah seluruh huruf dalam ulasan menjadi huruf kecil, sehingga jika ada kata yang sama tetapi memiliki perbedaan dalam penggunaan huruf kapital tidak akan dianggap berbeda [26].
- Filtering,**
Tahapan ini digunakan untuk menghapus angka, tanda simbol dan url yang tidak memiliki nilai dan hanya menyisakan kata-kata penting yang bermakna untuk proses analisa sentimen dari ulasan yang digunakan [19].
- Stopword Removal,**
Pada sebuah kalimat, terdapat kata-kata umum (contoh: “yang”, “di”, “dari”, “dan”, dan sebagainya) yang tidak memiliki makna, sehingga perlu dilakukan pembersihan dengan menggunakan kamus *stopword* bahasa Indonesia dalam *library* Sastrawi.
- Stemming,**
Penggunaan imbuhan dalam sebuah kata, tidak memiliki makna untuk dilakukan proses analisa sentimen [27], sehingga perlu adanya proses mengembalikan ke bentuk dasar menggunakan *library* Sastrawi.

Tabel 1. Representasi Pra-Pemrosesan untuk Naïve Bayes

Teks Asli: Semenjak di upgrade.. SHOPEE JADI LEMOT	
Tahap 1: Case Folding	semenjak di upgrade.. shopee jadi lemot
Tahap 2: Filtering	semenjak di upgrade shopee jadi lemot
Tahap 3: Stopword Removal	semenjak upgrade shopee jadi lemot
Tahap 4: Stemming	semenjak upgrade shopee jadi lot

3.4. Pra-Pemrosesan untuk IndoBERT

Berbeda dengan Naïve Bayes, algoritma IndoBERT memiliki tingkat sensitif yang tinggi terhadap teks dan menggunakan *tokenizing* bawaan. Oleh karena itu, tahapan pra-pemrosesan tidak seketat Naïve Bayes agar tidak menghilangkan banyak informasi kontekstual dari teks. Tahapan pra-pemrosesan untuk IndoBERT hanya terdiri dari 2 (dua) tahapan, yaitu:

- Filtering** (pembersihan url/emoji),
Tahapan ini hanya berfokus pada penghapusan URL dan emoji yang tidak relevan dengan sentimen dari teks. Tanda baca, penggunaan huruf kapital, dan imbuhan masih dipertahankan karena IndoBERT dapat memahami informasi struktural dan semantik yang terdapat pada bentuk teks.
- Tokenizing IndoBERT,**
Pada tahapan ini, teks diubah menjadi sebuah token menggunakan *tokenizer* IndoBERT [28]. Proses ini menggunakan pendekatan *subword*, sehingga sebuah kata dapat direpresentasikan sebagai satu atau beberapa token.

Tabel 2. Representasi Pra-Pemrosesan untuk IndoBERT

Teks Asli: Semenjak di upgrade.. SHOPEE JADI LEMOT	
Tahap 1: Filtering	Semenjak di upgrade.. SHOPEE JADI LEMOT
Tahap 2: Tokenizing IndoBERT	[CLS] Semenjak di upgrade.. SHOPEE JADI LEMOT [SEP]

Data yang sudah dilakukan tahapan pra-pemrosesan pada kedua algoritma, kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20 menggunakan fungsi **train_test_split**. Penggunaan parameter *random_state=42* agar hasil dari proses pengacakan data tetap konsisten, sehingga setiap kali proses pembagian data dilakukan, data latih dan data uji tetap sama.

3.5. Ekstraksi Fitur TF-IDF untuk Naïve Bayes

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) berfungsi untuk mengubah data teks menjadi representasi matriks numerik [29] agar dapat digunakan oleh algoritma Naïve Bayes. Terdapat 2 (dua) konsep dalam metode ini, yaitu *Term Frequency* (TF) untuk menghitung frekuensi kemunculan kata dalam suatu dokumen dan *Inverse Document Frequency* (IDF) untuk mengukur seberapa sering atau tidak kata tersebut muncul pada seluruh dokumen. Hasil akhir adalah bobot numerik yang membentuk matriks fitur.

3.6. Pelatihan Model Naïve Bayes

Pada tahapan ini, algoritma akan mengenali pola hubungan antara teks dengan skor sentimen target. Objek diinisialisasi menggunakan *library Multinomial Naïve Bayes* melalui fungsi *MultinomialNB()*. Pemilihan fungsi tersebut karena data berupa bobot TF-IDF dari dokumen teks.

Selanjutnya, model dilatih menggunakan fungsi *nb_model.fit(X_train_tfidf, y_train)*. Selama proses berlangsung, algoritma Naïve Bayes akan menghitung parameter probabilitas statistik, yang meliputi probabilitas di masing-masing kelas sentimen dengan probabilitas kemunculan setiap kata di kelas tertentu.

3.7. Fine-Tuning IndoBERT

Tahap *fine-tuning* digunakan untuk menyesuaikan algoritma IndoBERT agar dapat mengenali klasifikasi sentiment pada data ulasan. Proses ini dibagi ke dalam beberapa tahapan konfigurasi, yaitu:

- Inisialisasi Model Klasifikasi,
Fungsi yang digunakan untuk model klasifikasi adalah `AutoModelForSequenceClassification`. `from_pretrained(model_name, num_labels=2)` dari *library Hugging Face*. Fungsi ini ditambahkan dengan nilai `num_labels=2`, yang merepresentasikan bahwa terdapat dua kelas target sentimen, yaitu positif dan negatif.
- Konfigurasi Parameter Pelatihan,
Konfigurasi dilakukan melalui kelas `TrainingArguments` dengan menentukan direktori penyimpanan, jumlah 3 (tiga) epoch untuk perulangan pelatihan, dan ukuran batch data per *device* sebanyak 16. Selain itu, pada `eval_strategy` dan `save_strategy` diatur berdasarkan akhir dari epoch.
- Inisialisasi *Trainer*,
Seluruh komponen pelatihan akan digabung ke dalam sebuah objek bernama *Trainer* yang digunakan untuk mengelola seluruh jalannya proses pelatihan secara otomatis dengan memasukkan model utama, parameter pelatihan, dan data latih tokenisasi. Tujuan dari objek tersebut adalah memantau peningkatan kualitas model pada setiap akhir iterasi epoch.

4. HASIL DAN PEMBAHASAN

Pengujian performa untuk algoritma Naïve Bayes dengan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma IndoBERT dilakukan untuk mengevaluasi analisa sentimen pada ulasan pengguna. Evaluasi model diukur menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score* dengan divisualisasikan melalui *confusion matrix*.

4.1. Evaluasi Algoritma Naïve Bayes

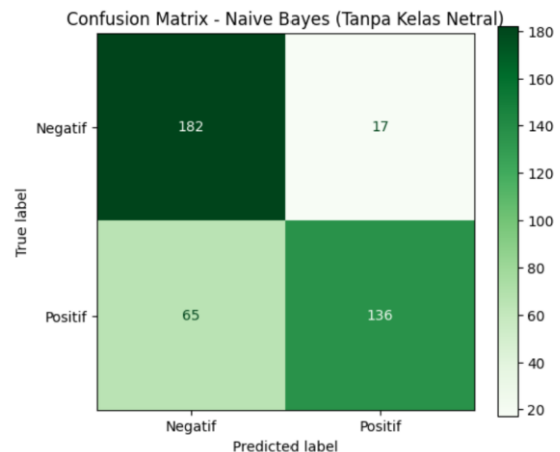
Hasil mengujikan algoritma Naïve Bayes berbasis TF-IDF mencapai akurasi sebesar 80% yang ditunjukkan pada tabel 3.

Tabel 3. Hasil performa pengujian dari algoritma Naïve Bayes.

	precision	recall	F1-score	support
Negatif	0.74	0.91	0.82	199
Positif	0.89	0.68	0.77	201
accuracy			0.80	400
macro avg	0.81	0.80	0.79	400
weighted avg	0.81	0.80	0.79	400

Distribusi ketepatan prediksi algoritma Naïve Bayes dengan TF-IDF terhadap masing-masing kelas

aktual berdasarkan *confusion matrix* digambarkan pada Gambar 3.



Gambar 3. Hasil *Confussion Matrix* dari Naïve Bayes.

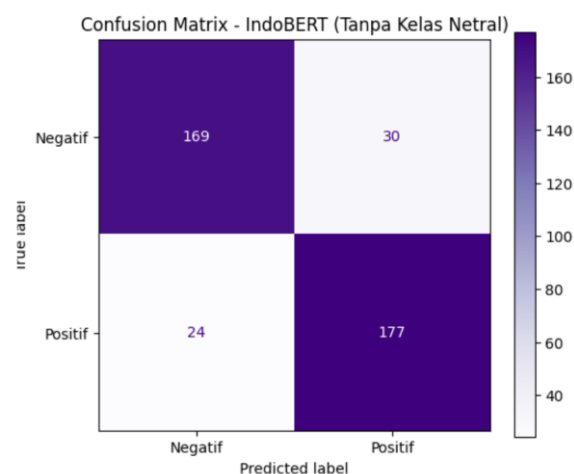
4.2. Evaluasi Algoritma IndoBERT

Pengujian berikutnya menggunakan algoritma IndoBERT dengan konfigurasi 3 (tiga) epoch menunjukkan peningkatan performa dengan akurasi mencapai 86.5% pada dataset uji coba yang sama. Rincian hasil pengujian ditampilkan ke dalam tabel 4.

Tabel 4. Hasil performa pengujian dari algoritma IndoBERT.

	precision	recall	F1-score	support
Negatif	0.88	0.85	0.86	199
Positif	0.86	0.88	0.87	201
accuracy			0.865	400
macro avg	0.86	0.86	0.86	400
weighted avg	0.86	0.86	0.86	400

Distribusi ketepatan dari algoritma IndoBERT berdasarkan hasil *confusion matrix* ditunjukkan pada gambar 4.



Gambar 4. Hasil *Confussion Matrix* dari IndoBERT.

4.3. Analisa Perbedaan Hasil

Algoritma Naïve Bayes dengan TF-IDF menggunakan perhitungan frekuensi kemunculan kata dan bobot setiap dokumen. Pendekatan ini cukup efektif untuk menangkap kata kunci yang spesifik pada ulasan negatif dengan hasil *recall* negative yang tinggi sebesar 0.91. Namun, TF-IDF mengabaikan urutan kata dan konteks semantic, sehingga menimbulkan kesalahan pada kata yang ambigu. Berbanding terbalik dengan IndoBERT yang unggul dengan akurasi 86.5% karena kemampuannya dalam memahami konteks kalimat secara utuh.

Pada Naïve Bayes, terdapat ketimpangan performa yang mana kelas positif memiliki *recall* yang rendah sebesar 0.68, yang mengakibatkan tingginya angka *false negative* yang mencapai 65 ulasan positif yang keliru dipresiksi sebagai ulasan negatif. Di lain sisi, algoritma IndoBERT dapat mengatasi hal tersebut dengan mendistribusikan performa secara lebih merata, angka *false negative* hanya 24 ulasan yang keliru diprediksi sebagai ulasan negatif dan nilai *recall* kelas positif sebesar 0.88.

Dari kedua analisa diatas, IndoBERT masih lebih unggul dibandingkan Naïve Bayes. Namun, IndoBERT menghasilkan *false positive* pada kelas negative sebanyak 30 ulasan, sedangkan Naïve Bayes hanya 17 ulasan. Kondisi ini menunjukkan bahwa masih terdapat ulasan negatif yang mengandung beberapa kata dengan karakteristik atau makna yang menyerupai pujian maupun sentimen positif, sehingga ulasan tersebut belum dapat diklasifikasikan secara tepat oleh IndoBERT.

5. KESIMPULAN DAN SARAN

Penelitian ini menggunakan 2 (dua) algoritma yang berbeda yaitu Naïve Bayes dengan TF-IDF dan IndoBERT. Data yang digunakan dari ulasan pengguna pada sebuah *e-commerce* yaitu Shopee. Tahapan pra-pemrosesan teks disesuaikan dengan karakteristik dari algoritma, yang mana pada algoritma Naïve Bayes terdapat banyak tahapan untuk pembersihan data, sedangkan algoritma IndoBERT hanya 2 (dua) proses saja untuk pembersihan data.

Algoritma Naïve Bayes dengan representasi fitur TF-IDF mampu mencapai tingkat akurasi sebesar 80% dengan kemampuannya dalam memprediksi ulasan negatif, meskipun memiliki tingkat kesalahan yang besar dalam memprediksi ulasan positif. Sementara itu, IndoBERT dengan proses *fine-tuning* jauh lebih unggul dengan akurasi sebesar 86,5% dan skor F1 sebesar 0.867. Hal ini membuktikan bahwa arsitektur *transformer* lebih efektif dalam memahami konteks semantic dan struktur bahasa informal seperti ulasan pelanggan.

Pada awal penelitian, terdapat 3 (tiga) klasifikasi sentimen, yaitu positif, negatif, dan netral. Namun penelitian ini hanya menggunakan 2 (dua) kelas saja dengan menghilangkan kelas netral. Oleh karena itu, saran untuk peneliti selanjutnya adalah penggunaan kelas netral untuk memprediksi sebuah ulasan

pelanggan dengan menambah jumlah epoch pada IndoBERT.

DAFTAR PUSTAKA

- [1] N. Nabila, T. W. Dari, and J. Hendra K, "Perubahan Pola dan Perilaku Konsumsi Masyarakat di Tengah Digitalisasi Pasca Pandemi Covid-19," *JAMPARING J. Akunt. Manaj. Pariwisata dan Pembelajaran Konseling*, vol. 3, no. 1, pp. 147–159, 2025, doi: 10.57235/jamparing.v3i1.4529.
- [2] D. Sulisty, F. Ahda, and V. A. Fitria, "Epistemologi dalam Natural Language Processing," *J. Inov. Teknol. dan Edukasi Tek.*, vol. 1, no. 9, pp. 652–664, 2021, doi: 10.17977/um068v1i92021p652-664.
- [3] D. Yogi Pratama, A. Saputra, Z. Saifullah, R. Dwiputra, and D. Apriadi, "Analisis Sentimen Kepuasan Pengguna Aplikasi Shopee Berdasarkan Ulasan Google Play Store Menggunakan Metode Naive Bayes," *Digit. Transform. Technol.*, vol. 6, no. 1, pp. 341–350, 2026, [Online]. Available: <https://doi.org/10.47709/digitech.v6i1.8351>
- [4] W. Alfiani, D. Abdul Fatah, and F. Irahmani, "Penerapan Algoritma Naïve Bayes Untuk Analisis Sentimen Pada Media Sosial X Terhadap Performa Tim Nasional Sepak Bola Indonesia Di Era Kepemimpinan Shin Tae-Yong," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 3, pp. 3969–3977, 2025, doi: 10.36040/jati.v9i3.13451.
- [5] A. Sukma Wanda and Usman, "Implementation of Naïve Bayes for Tokopedia Review Sentiment Classification Based on Web Dashboard," *J. Armada Inform.*, vol. 10, no. 1, pp. 103–113, 2026, doi: 10.36520/jai.v10i1.245.
- [6] N. F. Adhim and N. Cahyono, "Optimization of IndoBERT for Sentiment Analysis of FOMO on Social Media Through Fine-Tuning and Hybrid Labeling," *J. Appl. Informatics Comput.*, vol. 9, no. 6, pp. 3786–3797, 2025, doi: 10.30871/jaic.v9i6.11686.
- [7] D. K. Sumartha, "Opini Publik Terhadap Kebijakan Kenaikan Ukt Di Era Pemerintahan," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 3, 2025.
- [8] M. Rafli and B. Assegaf, "PENERAPAN INDOBERT UNTUK ANALISIS SENTIMEN BERBASIS ASPEK PADA EVALUASI APLIKASI LAYANAN PUBLIK," vol. 12, no. 1, pp. 93–101, 2026.
- [9] F. Rafiandi Andhika, W. Witanti, and P. N. Sabrina, "Analisis Sentimen Menggunakan Metode IndoBERT pada Ulasan Aplikasi Zoom Menggunakan Fitur Ekstrasi GloVe," *METIK J.*, vol. 9, pp. 439–448, 2025, doi: 10.47002/metik.v9i2.1098.
- [10] T. Dzulkarnain, D. E. Ratnawati, and B. Rahayudi, "Penggunaan Metode Naïve Bayes

- Classifier pada Analisis Sentimen Penilaian Masyarakat Terhadap Pelayanan Rumah Sakit di Malang,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 993–1000, 2024, doi: 10.25126/jtiik.2024117979.
- [11] I. G. S. D. Putra and I. N. T. A. Putra, “Implementasi Metode Naïve Bayes Pada Analisis Sentimen Pengguna Aplikasi Mobile Kita Bisa,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6423.
- [12] F. Ariany, A. Tri Prastowo, I. Ahmad, and Q. Jafar Adrian, “Algoritma Naïve Bayes dalam Analisis Komperatif Sentimen Dompot Digital,” *J. Tekno Kompak*, vol. 20, no. 1, pp. 364–370, 2026, doi: 10.33365/jtk.v20i1.1454.
- [13] N. Nurwanda, N. Suarna, and W. Prihartono, “Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1841–1846, 2024, doi: 10.36040/jati.v8i2.8469.
- [14] R. P. K. Nugroho, C. M. Joy Sihombing, and F. Ihda Husni Husodo, “Meninjau Perkembangan Teknologi NLP dan LLM dalam Meningkatkan Kualitas Percakapan serta Pemahaman Konteks oleh Chatbot,” *J. Profesi Ins. Univ. Lampung*, vol. 6, no. 2, pp. 2722–5771, 2025, [Online]. Available: <https://jpi.eng.unila.ac.id/index.php/ojs/article/view/274>
- [15] S. . Nasution, *Buku Monograf*. 2022. [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI
- [16] H. Wijaya and N. Hayati, “Natural Language Processing (Nlp) Untuk Analisis Sentimen Ulasan Seblak Bandung Pedas Kudus,” *J. Bus. Audit Inf. Syst.*, vol. 8, no. 1, pp. 13–22, 2025, [Online]. Available: <http://journal.ubm.ac.id/index.php/jbase>
- [17] Rizal Chandra Rivaldi and T.D. Wismarini, “Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP),” *Elkom J. Elektron. dan Komput.*, vol. 17, no. 1, pp. 120–128, 2024, doi: 10.51903/elkom.v17i1.1680.
- [18] M. Bagus Dikal Putra and E. Setiawan, “Metode Lexicon Based Untuk Analisis Sentimen Pengguna Twitter Terhadap Kinerja Isp,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 6, pp. 12079–12087, 2024, doi: 10.36040/jati.v8i6.11765.
- [19] D. K. Wardani, I. Syarif, and T. Badriyah, “Detecting Alter Ego Accounts using Social Media Mining,” *J. RESTI*, vol. 7, no. 3, pp. 602–608, 2023, doi: 10.29207/resti.v7i3.4919.
- [20] B. A. Santoso, B. I. Nugroho, and D. U. Asyfiya, “Perbandingan Algoritma Naïve Bayes, Support Vector Machine, dan Random Forest Untuk Analisis Sentimen Komentar Politik Youtube,” *TIN Terap. Inform. Nusantara*, vol. 6, no. 4, pp. 391–400, 2025, doi: 10.47065/tin.v6i4.8326.
- [21] A. Ramadhan Firdaus, F. M. Ayudewi, and A. Firdonsyah, “Analisis Sentimen Multi-Platform Digital Menggunakan Bert Dengan Pendekatan Penyeimbangan Distribusi Kelas,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 10, no. 1, pp. 1220–1226, 2026, doi: 10.36040/jati.v10i1.16946.
- [22] I. W. Bagus Satria Putra Pratama, M. Adi Paramartha Putra, and A. A. Istri Ita Paramitha, “Deteksi Berita Hoaks Bahasa Indonesia Menggunakan Model Hybrid Indobert-Bilstm Dengan Teknik Hyperparameter Tuning,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 10, no. 2, pp. 2830–2838, 2026, doi: 10.36040/jati.v10i2.17872.
- [23] D. Q. A. S. Covid, R. A. Pramana, and A. Romadhony, “Identifikasi Similar Question dengan IndoBERT (Studi Kasus),” 2, vol. 2, no. 1, pp. 12–17, 2024.
- [24] Dina Wulan Yekti rahayu, Khothibul Umam, and Maya Rini Handayani, “Performance of Machine Learning Algorithms on Imbalanced Sentiment Datasets Without Balancing Techniques,” *J. Appl. Informatics Comput.*, vol. 9, no. 3, pp. 998–1005, 2025, doi: 10.30871/jaic.v9i3.9584.
- [25] A. D. Fitriyanto and P. Purwanto, “Analisis Sentimen Ulasan DANA Dari Play Store dengan Metode SVM, Logistic Regression, Naive Bayes dan KNN,” *Build. Informatics, Technol. Sci.*, vol. 7, no. 3, pp. 1887–1899, 2025, doi: 10.47065/bits.v7i3.8769.
- [26] P. Studi Informatika, F. Matematika dan Ilmu Pengetahuan Alam, J. Raya Kampus UNUD, B. Jimbaran, and K. Selatan, “Analisis Sentimen Pengguna Terhadap Ulasan Aplikasi Duolingo Menggunakan Metode Logistic Regression Made Dinda Radityaswari a1 , Luh Gede Astuti a2,” *Jnatia*, vol. 4, no. 1, pp. 53–62, 2025.
- [27] N. Anisa, “Penerapan Lexicon Based Untuk Analisis Sentimen pada Game PUBG dengan Ekstraksi Fitur TF-IDF,” *J. Pract. Comput. Sci.*, vol. 5, no. 1, pp. 35–41, 2025, doi: 10.37366/jpcs.v5i1.5826.
- [28] A. N. Wahyuni *et al.*, “Implementasi Model Indobert Dan Mbert,” vol. 10, no. 2, pp. 2736–2743, 2026.
- [29] W. Ciko Nanda Putra and M. M. Mulki, “Klasifikasi Kategori Berita Berdasarkan Judul Menggunakan TF-IDF dan Support Vector Machine,” *Semin. Nas. Teknol. Sains*, vol. 5, pp. 449–456, 2026.