

Analisis Sentimen Pada Ulasan Tempat Wisata Desa Ponggok Menggunakan *Natural Language Processing*

Anggun Berlian Agustina^{1*}, Hanifah Permatasari², Joni Maulindar³

¹Teknik Informatika/Ilmu Komputer
Universitas Duta Bangsa Surakarta
¹*210103178@mhs.udb.ac.id

²Teknik Informatika/Ilmu Komputer
Universitas Duta Bangsa Surakarta
²hanifah_permatasari@udb.ac.id

³Teknik Informatika/Ilmu Komputer
Universitas Duta Bangsa Surakarta
³joni_maulindar@udb.ac.id

Abstrak— Ulasan pengunjung di platform digital telah menjadi faktor krusial yang mempengaruhi reputasi destinasi wisata seperti Desa Ponggok. Namun, volume data ulasan yang masif dan bersifat informal menyulitkan analisis manual, sehingga diperlukan pendekatan komputasional untuk mengekstrak wawasan sentimen secara efektif. Penelitian ini bertujuan untuk merancang dan mengimplementasikan model analisis sentimen pada ulasan berbahasa Indonesia mengenai tempat wisata Desa Ponggok. Metode yang digunakan adalah *fine-tuning* model *pre-trained* berbasis *Transformer*, yaitu IndoBERT, yang unggul dalam pemahaman konteks bahasa alami. Proses penelitian mencakup pengumpulan data dari Google Reviews, pra-pemrosesan data termasuk penanganan class imbalance dengan *RandomOverSampler*, pelatihan model, dan evaluasi menggunakan *confusion matrix*, akurasi, presisi, *recall*, serta F1-score. Hasil evaluasi pada data uji menunjukkan akurasi keseluruhan sebesar 79% dan F1-score 0.87 untuk kelas positif, namun menghadapi tantangan pada kelas netral dan negatif. Model yang telah dilatih kemudian berhasil diimplementasikan ke dalam sebuah aplikasi web interaktif menggunakan framework *Streamlit*, yang memungkinkan para pemangku kepentingan untuk melakukan inferensi sentimen secara langsung dan memvalidasi kegunaan praktis dari model.

Kata kunci— Analisis Sentimen, Natural Language Processing, IndoBERT, Ulasan Wisata, Streamlit

Abstract— Visitor reviews on digital platforms have become a crucial factor influencing the reputation of tourist destinations like Desa Ponggok. However, the massive volume and informal nature of review data make manual analysis inefficient, necessitating a computational approach to effectively extract sentiment insights. This research aims to design and implement a sentiment analysis model for Indonesian-language reviews of the Desa Ponggok tourist destination. The methodology employed is fine-tuning a pre-trained Transformer-based model, IndoBERT, which excels in understanding natural language context. The research process included data collection from Google Reviews, data preprocessing including handling class imbalance with *RandomOverSampler*, model training, and evaluation using confusion matrix, accuracy, precision, recall, and F1-score. The evaluation on the test set yielded an overall accuracy of 79% and an F1-score of 0.87 for the positive class, while facing challenges with the neutral and negative classes. The trained model was successfully deployed into an interactive web application using the *Streamlit* framework, enabling stakeholders to perform sentiment inference directly and validate the model's practical utility.

Keywords— Sentiment Analysis, Natural Language Processing, IndoBERT, Tourism Reviews, Streamlit

I. PENDAHULUAN

Desa Ponggok merupakan salah satu destinasi wisata unggulan di Indonesia yang terkenal dengan keindahan alamnya, terutama kolam renang alami dan pemandangan bawah airnya yang menakjubkan. Seiring dengan perkembangan teknologi dan media sosial, ulasan pengunjung tentang tempat wisata ini semakin banyak tersebar di platform online seperti Google Reviews, Trip Advisor, dan media sosial lainnya [1]. Ulasan-ulasan ini mengandung informasi berharga yang dapat digunakan untuk memahami kepuasan pengunjung serta mengidentifikasi area yang perlu ditingkatkan. Di era mudahnya informasi, ulasan menjadi salah satu tolak ukur bagi pengunjung sebelum mengunjungi suatu tempat [2]. Terutama alam bidang pelayanan,

pengunjung cenderung lebih memperhatikan kualitas ulasan karena layanan tersebut tidak bisa dinilai sebelum benar-benar digunakan [3].

Namun, volume dan keragaman ulasan yang masif membuat analisis manual menjadi tidak efisien, sehingga memerlukan pendekatan komputasional seperti *Natural Language Processing* (NLP) untuk ekstraksi wawasan secara otomatis. Namun hadirnya NLP bukan tanpa hambatan, elemen seperti sarkasme, ironi, negasi, disambiguasi arti kata, dan ambiguitas [4] dapat mempengaruhi *output* dari model secara signifikan. Seiring hadirnya model-model yang lebih kuat seperti GPT, BERT, dan sejenisnya dapat membuka potensi baru yang akan lebih kuat [5]. Penelitian ini bertujuan untuk mengimplementasikan teknik *NLP* dalam menganalisis sentimen ulasan pengunjung Desa

Ponggok. Dengan demikian, diharapkan dapat diperoleh wawasan yang akurat dan membantu pengelola wisata Desa Ponggok untuk mengambil keputusan berbasis data guna meningkatkan kualitas pelayanan dan pengalaman pengunjung.

1.1 Penelitian Terkait

Analisis sentimen pada ulasan pengguna telah menjadi area penelitian yang aktif, terutama pada domain evaluasi produk dan layanan digital. Sebagai contoh, penelitian pada ulasan *smartphone* di Instagram, [6] menggunakan *Naive Bayes* dan berhasil mencapai akurasi 81% dengan varian *Multinomial*. Sedangkan di sektor teknologi finansial, [7] menerapkan metode *Random Forest* pada ulasan aplikasi DANA, yang terbukti efektif dengan capaian 84% untuk semua metrik utama (akurasi, presisi, *recall*, dan F1-Score). Penelitian lain di bidang layanan kesehatan oleh [8] menunjukkan keunggulan metode *K-Nearest Neighbors (KNN)* yang menghasilkan akurasi 95% dan AUC 0,985 saat menganalisis ulasan aplikasi Halodoc, melampaui performa *Naive Bayes* dan *SVM*.

Meskipun penelitian-penelitian tersebut berhasil menunjukkan kapabilitas analisis sentimen, terdapat dua celah utama yang coba diatasi oleh riset ini. Pertama, seluruh studi kasus diatas menggunakan pendekatan menggunakan metode *machine learning* klasik yang dimana memiliki keterbatasan untuk memahami konteks secara keseluruhan, sehingga apabila kalimat terpotong atau tidak lengkap maka dapat menghasilkan *output* yang tidak sesuai. Seperti studi yang dilakukan oleh [9] dengan membandingkan 15 model *machine learning* klasik dengan model *deep learning* seperti BERT, RoBERTa, dan 3 varian dari FinBERT, hasilnya menunjukkan model *state-of-art* berbasis *transformer* mengungguli semua model *machine learning* klasik. Kedua, fokus domainnya cenderung pada produk dan aplikasi digital, sedangkan pada kasus ulasan tempat wisata yang memiliki karakteristik linguistik terkait nuansa, pengalaman, dan interaksi pelayanan

Penelitian ini memiliki perbedaan mendasar dari studi-studi di atas. Perbedaan utama terletak pada domain penelitian, yaitu ulasan destinasi

wisata yang khas, serta metodologi yang digunakan. Melihat keterbatasan metode klasik tersebut dalam memahami konteks kalimat, penelitian ini menerapkan pendekatan *state-of-the-art* dengan melakukan *fine-tuning* pada model berbasis *transformer* (IndoBERT) yang dirancang khusus untuk Bahasa Indonesia. Lebih lanjut, hasil dari model ini diimplementasikan menjadi sebuah sistem interaktif menggunakan bantuan *framework Streamlit*.

1.2 Landasan Teori

Penelitian ini berada dalam domain *natural language processing* yang dijelaskan oleh [10] sebagai cabang dari *Artificial Intelligence (AI)* yang berfokus pada interaksi antara manusia dan komputer melalui bahasa alami. NLP mempelajari berbagai aspek penggunaan bahasa, mulai dari kata, kalimat, paragraf, hingga keseluruhan dokumen. Salah satu penerapan praktis yang relevan dengan NLP adalah analisis sentimen yang dijelaskan [11] dalam jurnalnya yang berjudul "*A Survey on Sentiment Analysis Methods, Applications, and Challenges*" bahwa analisis sentiment bertujuan untuk mengidentifikasi sikap, emosi, atau sentimen penulis terhadap suatu topik, entitas, atau produk.

Untuk menjalankan analisis sentimen secara akurat pada teks yang lebih skenario yang kompleks seperti sarkasme atau konteks yang tidak lengkap, penelitian modern beralih dari metode statistik ke model jaringan saraf tiruan atau *deep learning*. Sebuah terobosan fundamental dalam bidang ini adalah BERT (*Bidirectional Encoder Representations from Transformers*) yang diperkenalkan oleh *Google AI Language* pada akhir tahun 2018. Model ini dirancang untuk pembelajaran awal yang mendalam dalam merepresentasikan teks secara dua arah (*bidirectional*), sehingga dapat digunakan lebih lanjut dalam model pembelajaran mesin [12]. Keunggulan revolusioner inilah yang membuat BERT mampu menangkap konteks lebih baik dibandingkan dengan model tradisional yang digunakan pada penelitian-penelitian sebelumnya.

Mengingat pada penelitian berfokus pada bahasa Indonesia, maka penggunaan model turunan BERT yang dirancang spesifik untuk

bahasa Indonesia akan lebih baik, yaitu model IndoBERT. Pengembangan model IndoBERT sendiri didasarkan oleh kurangnya bahasa Indonesia dalam penelitian NLP meskipun bahasa Indonesia digunakan oleh hampir 200 juta orang dan termasuk dalam 10 bahasa paling banyak digunakan di dunia. IndoBERT merupakan model berbasis transformer seperti BERT, yang dilatih secara khusus sebagai *masked language model* menggunakan *framework Hugging Face*, dengan konfigurasi bawaan BERT-Base [13].

II. METODOLOGI PENELITIAN

Pada proses pengembangan sistem, pendekatan yang digunakan bersifat iteratif, artinya pengembangan dilakukan secara bertahap dan berulang. Setiap iterasi melibatkan evaluasi terhadap hasil sementara, yang kemudian digunakan sebagai dasar untuk perbaikan atau penambahan fitur pada iterasi berikutnya hingga mendapatkan hasil yang di inginkan.

2.1 Data Collecting

Tahap pertama dalam penelitian ini adalah pengumpulan data ulasan pengunjung tentang tempat wisata Desa Ponggok. Data ulasan akan diambil dari berbagai platform online seperti Google Reviews, dan media sosial. Pengumpulan data dilakukan dengan menggunakan teknik web scraping dan akan dilakukan proses Extract, Transform, Load. Data yang dikumpulkan akan disimpan dalam format terstruktur seperti CSV atau JSON, dengan kolom-kolom seperti ulasan, rating, dan tanggal. Tahap ini bertujuan untuk mendapatkan dataset yang representatif dan cukup besar untuk digunakan dalam pelatihan model.

2.2 Preprocessing

Setelah data ulasan berhasil dikumpulkan, tahap selanjutnya adalah *preprocessing* data. Pada tahap ini, data teks akan dibersihkan dan dipersiapkan agar dapat diproses oleh model. Beberapa langkah *preprocessing* yang dilakukan meliputi normalisasi teks (mengubah kata tidak baku menjadi baku, menghilangkan karakter berulang, dan mengatasi singkatan), tokenisasi (memisahkan teks menjadi kata-kata individual).

Namun pada pendekatan menggunakan model transformers seperti IndoBert, *stemming* ataupun *lemmatization* tidak dilakukan karena model ini di desain untuk belajar dari konteks atau hubungan kata, penggunaan *stemming* ataupun *lemmatization* malah berpotensi membuat model menjadi tidak akurat Selain itu, ulasan akan diberi label sentimen (positif, negatif, atau netral) menggunakan pendekatan *lexicon based* yaitu pendekatan yang menggunakan kamus kata berlabel sentiment untuk menentukan polaritas suatu teks, apakah negatif, positif, atau netral. Metode ini dipilih karena merupakan pendekatan yang efisien, dan umum digunakan dalam penelitian analisis berbasis sentimen.

Selanjutnya untuk menangani kelas yang tidak seimbang (*class imbalance*) akan dilakukan agar model tidak bias pada kelas tertentu. Tahapan *preprocessing* ini sangat penting untuk memastikan data siap digunakan dalam pelatihan model.

2.3 Fine-tuning model IndoBert

Setelah data siap, tahap berikutnya adalah *fine-tuning* model *IndoBERT*. Model ini dipilih karena kemampuannya menangani teks Bahasa Indonesia dengan baik serta efisiensinya dalam penggunaan sumber daya. Dataset yang telah diproses akan dibagi menjadi tiga bagian, yaitu data latih (*train*), data validasi (*validation*), dan data uji (*test*) dengan rasio pembagian data sesuai dengan karakteristik dataset.

2.4 Evaluasi

Setelah model selesai di-*fine-tune*, tahap selanjutnya adalah evaluasi performa model. Evaluasi dilakukan menggunakan data uji yang belum pernah dilihat oleh model selama pelatihan. Beberapa fungsi seperti *confusion matrix*, dan *classification report* serta beberapa matriks evaluasi yang dijabarkan pada point 2.4.1 akan digunakan untuk menganalisis performa model pada setiap kelas sentimen (positif, negatif, dan netral).

2.4.1. Matriks Evaluasi

Untuk mengukur performa model klasifikasi multi-kelas secara komprehensif, beberapa metrik evaluasi digunakan. Metrik-metrik ini diturunkan dari *Confusion Matrix*.

- a. **Confusion Matrix:** Merupakan sebuah tabel yang memvisualisasikan kinerja model dengan membandingkan hasil prediksi dengan label aktual. Dari matriks ini, kita dapat melihat secara detail jenis-jenis kesalahan yang dibuat oleh model. Komponen utamanya adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).
- b. **Accuracy (Akurasi):** Mengukur persentase prediksi yang benar dari keseluruhan data uji. Meskipun umum digunakan, akurasi bisa menjadi matrik yang kurang representatif jika dataset tidak seimbang (*imbalanced*).

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

- c. **Precision (Presisi):** Mengukur tingkat ketepatan dari prediksi positif. Dari semua ulasan yang diprediksi sebagai kelas tertentu (misal: 'positif'), berapa persen yang benar-benar 'positif'.

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

- d. **Recall (Daya Ingat):** Mengukur tingkat keberhasilan model dalam menemukan kembali sebuah kelas. Dari semua ulasan yang seharusnya 'positif', berapa persen yang berhasil diprediksi dengan benar oleh model.

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

- e. **F1-Score:** Merupakan rata-rata harmonik dari *Precision* dan *Recall*. Metrik ini memberikan gambaran yang lebih seimbang mengenai kinerja model, terutama pada dataset yang tidak seimbang, karena mempertimbangkan baik *False Positive* maupun *False Negative*.

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

2.5 Deployment

Tahap *deployment* diimplementasikan menggunakan metode *prototyping*. Menurut [14] metode *prototyping* adalah mengaktualisasikan dan mewujudkan sebuah konsep dengan maksud untuk menganalisis, dengan prototipe sebagai contoh desain yang lebih pendek dari produk akhir. Keuntungan utama metode ini adalah efisiensi waktu dan biaya. Selain itu, pendekatan

prototipe memungkinkan pengiriman produk lebih awal kepada pengguna untuk mendapatkan umpan balik secara cepat [15]

Pada tahapan ini akan dikembangkan antarmuka menggunakan framework Streamlit. Streamlit adalah framework Python sumber terbuka yang memungkinkan pengembang membuat aplikasi web interaktif dengan cepat dan menggunakan sedikit kode. Dirancang khusus untuk membantu ilmuwan data dan machine learning engineers, Streamlit mempermudah konversi skrip data menjadi aplikasi web yang dapat dibagikan dengan mudah [16].

Dengan demikian, hasil analisis sentimen dapat diakses oleh pemangku kepentingan sehingga dapat membantu untuk memahami informasi yang sekiranya berguna untuk pengembangan tempat wisata lebih lanjut, karena setiap keputusan akan didasarkan oleh fakta atau data.

2.6 Pengujian dan validasi sistem

Tahap terakhir adalah pengujian dan validasi sistem. Pada tahap ini, sistem akan diuji secara fungsional untuk memastikan bahwa semua komponen berfungsi dengan baik. Hasil analisis sentimen yang dihasilkan oleh sistem akan divalidasi secara manual dengan membandingkannya terhadap label sentimen yang sebenarnya. Jika antarmuka pengguna dikembangkan, umpan balik dari pengguna juga akan dikumpulkan untuk perbaikan lebih lanjut. Tahap ini bertujuan untuk memastikan bahwa sistem bekerja dengan akurat dan dapat diandalkan dalam memberikan hasil analisis sentimen.

III. HASIL DAN PEMBAHASAN

3.1 Data Collecting

Pengumpulan data ulasan dari tiga destinasi wisata di Desa Ponggok dengan fokus bahasa Indonesia dilakukan secara terprogram menggunakan perangkat lunak Apify, mengingat adanya keterbatasan pada Google Review API resmi. Proses ini dilanjutkan dengan tahap pemeriksaan data untuk mengetahui nilai yang hilang dan data duplikat. Rangkuman statistik

dari proses pengumpulan dan pemeriksaan data, disajikan secara rinci pada Tabel 1.

Tabel 1. Statistik Deskriptif Dataset

Deskripsi	Jumlah / Nilai
Total Data Awal	6.000 ulasan
Jumlah Sumber Lokasi Wisata	3
Lokasi Wisata	Umbul Ponggok, Umbul Besuki, dan Umbul Sigedhang
Jumlah Data per Lokasi	2.000 ulasan
Jumlah Nilai Hilang pada Fitur 'text'	2.859 ulasan
Jumlah Data Duplikat	Tidak ada

3.2 Preprocessing

Setelah dataset berhasil dimuat dan diperiksa kondisi dataset secara keseluruhan, pra-pemrosesan diterapkan pada dataset untuk memastikan data bersih dan siap digunakan. Pra-pemrosesan ini meliputi beberapa tindakan seperti.

3.2.1 Data Cleaning

Berdasarkan hasil pemeriksaan awal pada dataset, teridentifikasi adanya 2.859 nilai yang hilang pada kolom 'text'. Mengingat kolom 'text' merupakan fitur utama yang krusial untuk analisis sentimen, ulasan tanpa teks tidak memiliki nilai informatif dan tidak dapat diolah. Oleh karena itu, semua baris data yang tidak memiliki isi pada kolom 'text' dihapus. Setelah proses pembersihan ini, jumlah data valid yang siap untuk tahap selanjutnya berkurang menjadi 3.141 ulasan.

3.2.2 Pembersihan text

Langkah selanjutnya adalah pembersihan teks yang bertujuan untuk menghilangkan *noise* dan menyeragamkan format setiap ulasan sebelum diproses oleh model. Proses ini mencakup dua tindakan utama: (1) penghapusan semua simbol, tanda baca, angka, dan karakter non-alfanumerik lainnya, dan (2) mengubah seluruh teks menjadi format huruf kecil (*case folding*). Contoh transformasi sebuah ulasan sebelum dan sesudah proses pembersihan disajikan pada Tabel 2.

Tabel 2. Pembersihan Text

Tahapan	Text Ulasan
Teks Asli	Tempatnya BAGUS bgt!! 🍷 Airnya jernih & seger... recommended deh, HTM cuma 15rb.
Teks Bersih	Tempatnya bagus banget airnya jernih dan segar recommended deh htm cuma rb

3.2.3 Data Labelling

Setelah teks dibersihkan, tahap selanjutnya adalah pelabelan data untuk menentukan polaritas sentimen pada setiap ulasan. Penelitian ini menerapkan pendekatan *Lexicon-Based Sentiment Analysis* yang menggunakan kamus sentimen berbahasa Indonesia yang telah terverifikasi. Proses pelabelan dilakukan secara terprogram dengan alur kerja sebagai berikut:

1. Inisialisasi Kamus Sentimen

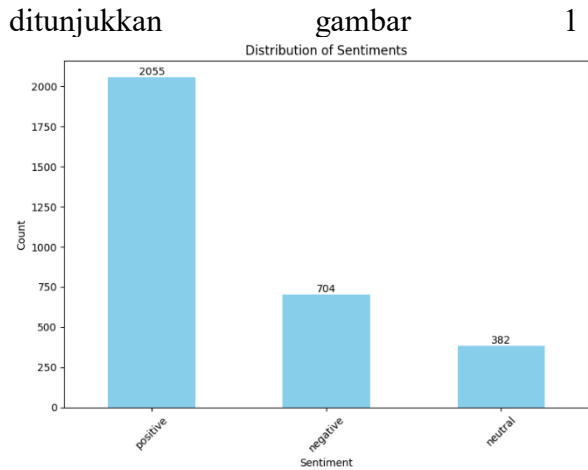
Sebuah kamus sentimen yang berisi daftar kata-kata positif dan negatif beserta skor polaritasnya dimuat. Kamus yang digunakan adalah kamus yang bersumber dari GitHub dengan *username* angelmetanosaa [17] di mana kata positif diberi skor +1 dan kata negatif diberi skor -1.

2. Perhitungan Skor

Untuk setiap ulasan, sistem memecah kalimat menjadi kata-kata. Setiap kata kemudian diperiksa keberadaannya di dalam kamus sentimen. Jika sebuah kata ditemukan, skor sentimen dari ulasan tersebut akan ditambahkan sesuai dengan skor kata tersebut di dalam kamus.

3. Penentuan Polaritas

Total skor dari sebuah ulasan kemudian digunakan untuk menentukan label sentimen akhirnya dengan aturan skor total lebih dari 0 diklasifikasikan sebagai positif, skor total kurang dari 0 diklasifikasikan sebagai negatif, skor total sama dengan 0 atau tidak ditemukan pada kamus diklasifikasikan sebagai netral. Setelah menerapkan proses pelabelan ini pada 3.141 baris data bersih, diperoleh distribusi kelas sentimen awal seperti yang



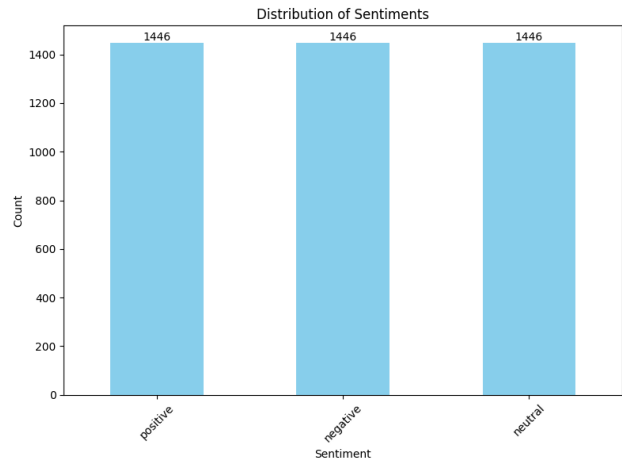
Gambar 1. Distribusi Sentimen Sebelum Penerapan Oversampling

Terdapat 2055 ulasan positif, 704 ulasan negatif, dan 382 ulasan netral. Berdasarkan distribusi diatas dapat disimpulkan bahwa adanya ketidakseimbangan antar kelas (*class imbalance*).

3.2.4 Data Preparation

Setelah pelabelan, langkah selanjutnya adalah mempersiapkan data untuk pelatihan model. Pertama, dataset yang terdiri dari 3.141 ulasan dibagi menjadi tiga set: data latih (70%), data validasi (15%), dan data uji (15%). Pembagian ini menghasilkan 2.199 data untuk latih, 471 data untuk validasi, dan 471 data untuk uji untuk memastikan model dapat dilatih dengan data yang cukup dan dilakukan validasi dalam waktu yang sama, lalu diuji menggunakan data yang belum pernah dilihat sebelumnya. Pemilihan dengan rasio diatas didasarkan jumlah data yang tidak terlalu banyak, sehingga model butuh proporsi yang seimbang.

Seperti yang ditunjukkan pada gambar 1, distribusi kelas pada dataset awal sangat tidak seimbang (*imbalanced*). Untuk mencegah model menjadi bias terhadap kelas mayoritas (positif), diterapkan teknik *oversampling* pada data latih menggunakan metode *RandomOverSampler*. Metode ini dipilih karena efektif untuk data tekstual dan tidak memerlukan transformasi ke bentuk numerik seperti SMOTE.



Gambar 2. Distribusi Sentimen Setelah Penerapan Oversampling

Pada gambar diatas terlihat bahwa masing-masing kelas sudah seimbang dengan jumlah 1446 baris data untuk setiap kelasnya. Dataset yang seimbang inilah yang kemudian digunakan untuk melatih model IndoBERT.

3.2.5 Tokenisasi (Tokenization)

Tahap akhir dari pra-pemrosesan adalah tokenisasi, yaitu proses mengubah sekuens teks menjadi urutan token numerik yang dapat dipahami oleh model. Penelitian ini menggunakan *tokenizer* bawaan dari model IndoBERT yang menerapkan algoritma WordPiece. Algoritma ini mampu menangani kata di luar kosakata (*out-of-vocabulary*) dengan memecahnya menjadi unit sub-kata. Contoh hasil proses tokenisasi disajikan pada Tabel 3.

Tabel 3. Tokenisasi Teks

Jenis	Teks
Sebelum tokenisasi	tempatnya bagus banget airnya jernih
Setelah tokenisasi	[[CLS], 'tempat', '##nya', 'bagus', 'banget', 'air', '##nya', 'jernih', '[SEP]]

3.3 Fine-tuning model IndoBert

Setelah dataset melalui tahapan pra-pemrosesan, maka siap digunakan untuk tahapan *modelling*. Setiap ulasan dan labelnya dibungkus dalam objek *Dataset()*, yang bertugas melakukan tokenisasi teks sesuai format IndoBERT dan mengubah label sentimen menjadi nilai numerik. Selanjutnya, *DataLoader()* digunakan untuk menyajikan data ke model dalam batch berukuran 32. Data latih diacak pada setiap epoch untuk

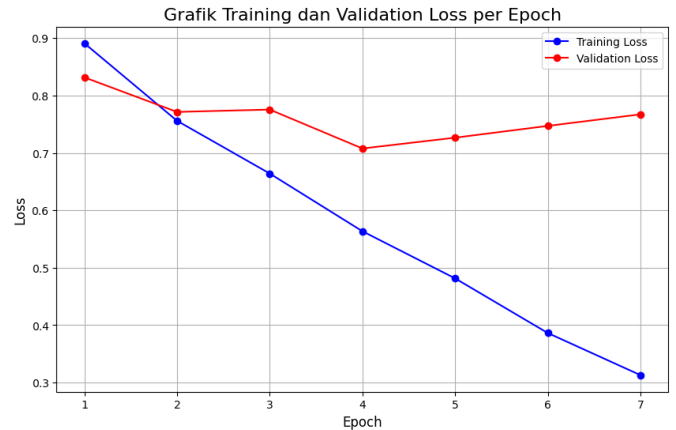
meningkatkan generalisasi, sementara urutan data validasi dan uji dijaga tetap untuk konsistensi evaluasi.

Proses optimisasi model menggunakan *optimizer Adam* dengan *learning rate* awal $3e-6$ dan *weight_decay* sebesar 0.01 untuk regularisasi. *Learning rate* ini diatur secara dinamis oleh *scheduler ReduceLROnPlateau*, yang akan menurunkannya jika *validation loss* tidak membaik selama 3 epoch berturut-turut. Proses fine-tuning dijalankan selama 7 epoch. Gambar 3 menunjukkan hasil pelatihan, di mana terlihat performa model pada data latih (TRAIN LOSS) dan data validasi (VALID LOSS) di setiap akhir epoch. Pada akhir pelatihan, model mencapai akurasi 0.97 pada data latih dan 0.80 pada data validasi, yang mengindikasikan adanya *overfitting*.

```
(Epoch 6) TRAIN LOSS:0.1025 ACC:0.96 F1:0.96 REC:0.96 PRE:0.96 LR:0.00000300
VALID LOSS:0.6000 ACC:0.79 F1:0.74 REC:0.72 PRE:0.77: 100% | 15/15 [00:04<00:00, 3.481t/s]
(Epoch 6) VALID LOSS:0.6000 ACC:0.79 F1:0.74 REC:0.72 PRE:0.77
(Epoch 7) TRAIN LOSS:0.1128 LR:0.00000300: 100% | 136/136 [01:42<00:00, 1.331t/s]
(Epoch 7) TRAIN LOSS:0.1128 ACC:0.97 F1:0.97 REC:0.97 PRE:0.97 LR:0.00000300
VALID LOSS:0.6482 ACC:0.80 F1:0.75 REC:0.73 PRE:0.79: 100% | 15/15 [00:04<00:00, 3.411t/s]
```

Gambar 3. Proses Training pada Model

Kemudian pelatihan divisualisasikan pada gambar 4, yang menampilkan perbandingan kurva *loss* untuk data latih dan data validasi. Dari grafik tersebut, dapat diamati bahwa kurva **training loss (biru)** menunjukkan penurunan yang konsisten di setiap *epoch*, mengindikasikan bahwa model berhasil mempelajari pola dari data latih. Sementara itu, kurva **validation loss (merah)** juga menunjukkan penurunan di epoch awal, namun mulai menunjukkan tren stagnan setelah epoch ke-3. Perbedaan yang semakin melebar antara kedua kurva ini merupakan indikasi kuat terjadinya *overfitting*, di mana model mulai kehilangan kemampuan memahami data yang belum pernah dilihat sebelumnya. Temuan visual mendukung kesimpulan bahwa diperlukan teknik regularisasi atau *hyperparameter tuning* lebih lanjut untuk mengatasi *overfitting*.



Gambar 4. Grafik Training and Validation Loss dari Model yang Sudah Dilatih

3.4 Evaluasi

Setelah proses *fine-tuning* selesai, tahap selanjutnya adalah evaluasi model. Tahap ini bertujuan untuk mengukur kinerja dan kapabilitas generalisasi dari model IndoBERT yang telah dilatih dalam mengklasifikasikan sentimen pada ulasan. Evaluasi dilakukan pada Data Uji (*Test Set*), yang merupakan data yang sama sekali baru dan belum pernah dilihat oleh model selama proses pelatihan maupun validasi.

Berdasarkan laporan klasifikasi pada gambar 5, pelatihan model mampu mendapatkan nilai yang hampir sempurna, yaitu 0.97 untuk keseluruhan matriks evaluasi meliputi akurasi, *f1-score*, *precision*, dan *recall* pada *training set*. Ini menunjukkan model dapat belajar dengan sangat baik dari dataset yang sudah disiapkan.

```
--- Classification Report ---
precision recall f1-score support
positive 0.87 0.86 0.87 316
neutral 0.71 0.59 0.65 51
negative 0.62 0.68 0.65 105

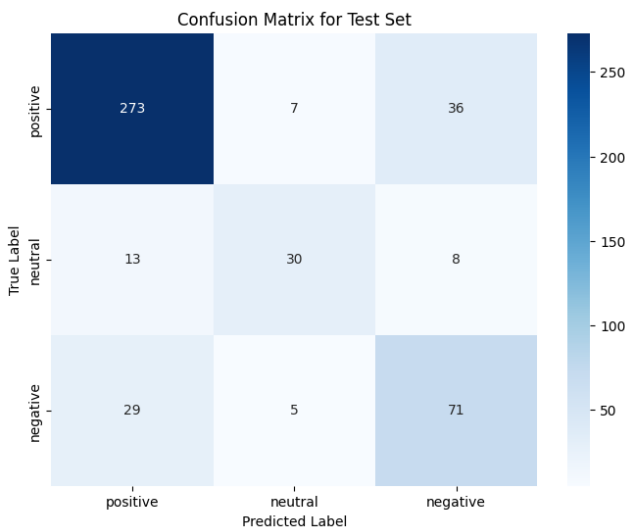
accuracy 0.79 472
macro avg 0.73 0.71 0.72 472
weighted avg 0.79 0.79 0.79 472
```

Gambar 5. Laporan Klasifikasi Hasil Evaluasi pada Data Uji

Lalu pada hasil evaluasi *testing set* menunjukkan akurasi keseluruhan sebesar 79%, yang mengindikasikan bahwa model secara umum mampu memberikan prediksi yang benar pada mayoritas data uji. Meskipun angka akurasi ini terlihat cukup baik, metrik ini bisa menjadi bias karena performa model yang sangat kuat pada kelas *positive* yang dominan dalam dataset. Analisis *F1-score* yang lebih mendalam

mengungkap bahwa di balik akurasi tersebut, kinerja model tidak merata: sangat andal pada kelas *positive* (*F1-score* 0.87), namun jauh lebih lemah pada kelas *neutral* dan *negative* (keduanya *F1-score* 0.65).

Untuk memahami penyebab kinerja yang tidak merata ini, analisis lebih dalam dilakukan menggunakan *Confusion Matrix*.



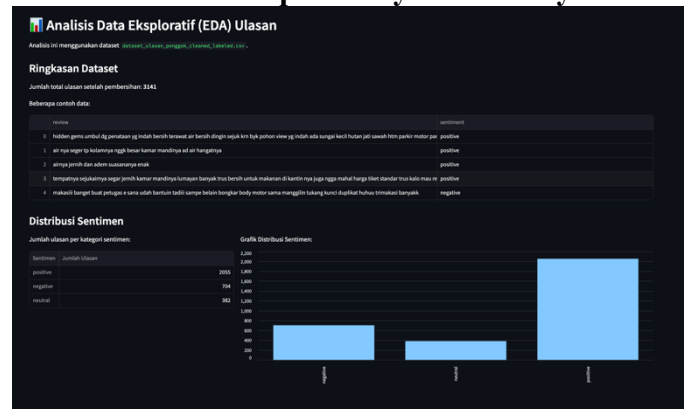
Gambar 6. Evaluasi Menggunakan Confusion Matrix pada Data Uji

Pada gambar 6 Terdapat 29 ulasan negatif yang salah diprediksi sebagai positif dan 36 ulasan positif yang salah diprediksi sebagai negatif. Ini adalah jenis kesalahan paling serius yang menunjukkan kesulitan model dalam memahami kalimat bernuansa atau yang mengandung negasi. Selain itu model juga kesulitan untuk pada kelas netral di mana 13 ulasan salah diklasifikasikan sebagai positif dan 8 sebagai negatif. Hal ini mengindikasikan ambiguitas pada ulasan netral yang seringkali mengandung kata-kata dari kedua polaritas. Rendahnya performa pada kelas negatif dan netral diduga disebabkan oleh bias leksikal dalam korpus data. Meskipun telah dilakukan penyeimbangan kelas (*oversampling*), distribusi kosakata secara inheren masih didominasi oleh penanda sentimen positif yang lebih sering muncul. Hal ini mengindikasikan bahwa penyeimbangan jumlah data saja tidak cukup jika tidak diiringi dengan diversifikasi kosakata pada kelas minoritas, terutama untuk menangkap variasi kalimat negatif dan netral yang ambigu.

3.5 Deployment

Sebagai tahap akhir penelitian, model analisis sentimen yang telah dikembangkan diimplementasikan menjadi sebuah aplikasi web interaktif menggunakan *framework* Streamlit. Aplikasi ini dirancang dengan dua fungsionalitas utama yang dapat diakses melalui halaman terpisah:

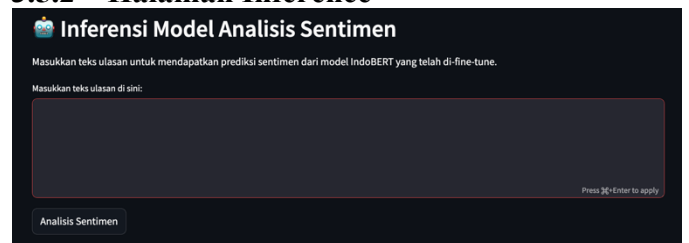
3.5.1 Halaman Exploratory Data Analysis



Gambar 7. Halaman Exploratory Data Analysis

Halaman ini berfungsi untuk menyajikan dataset yang telah dibersihkan kepada pengguna. Tujuannya adalah untuk memberikan transparansi mengenai data yang digunakan dalam melatih model. Fitur utama pada halaman ini adalah tabel data interaktif dan visualisasi distribusi sentimen (positif, netral, negatif) dalam bentuk diagram batang, yang memungkinkan pemahaman cepat mengenai komposisi data.

3.5.2 Halaman Inference



Gambar 8. Halaman Inference

Halaman ini merupakan antarmuka utama untuk menguji model secara langsung. Pengguna dapat memasukkan kalimat ulasan baru pada area teks yang tersedia. Setelah menekan tombol "Analisis Sentimen", sistem akan memproses input tersebut, melakukan inferensi menggunakan model yang telah dimuat, dan menampilkan hasil klasifikasi sentimen (Positif, Netral, atau Negatif) secara langsung pada layar.

3.6 Pengujian dan validasi system

Setelah model berhasil dimuat maka dilakukan pengujian sistem dengan memasukan contoh teks ulasan. Pengujian fungsional dilakukan untuk memvalidasi kinerja sistem pada berbagai skenario input teks. Beberapa kasus uji representatif dan hasilnya disajikan pada Tabel 4.

Tabel 4. Hasil Pengujian

No	Skenario	Text	Hasil Prediksi	Skor Keperca yaan	Hasil Validasi
1	Kalimat Positif	"Harganya murah, airnya seger pollllll, bersih lagi, bakalan sering balik kesini!!"	Positif	99.46%	Sesuai
2	Kalimat Negatif	"Tempatnya kotor, apa-apa harus bayar lagi, tempatnya juga ga tertata, nyesel bangettt kesini"	Negatif	70.36%	Sesuai
3	Kalimat Netral	"Pelayanan dan tempatnya standar la untuk harganya, kalau sempet aja mungkin kesini"	Netral	68.76%	Sesuai
4	Sarkasme	"Bagus banget tempatnya, saking bagusya sampai gak mau balik lagi deh."	Positif	85.12%	Tidak sesuai

WDari tabel pengujian, terlihat bahwa sistem mampu memprediksi dengan benar untuk kalimat dengan sentimen yang jelas (positif, negatif, netral). Namun, pada kasus uji batas seperti kalimat sarkastik model menunjukkan keterbatasannya dengan memprediksi sentimen secara harfiah (positif), yang tidak sesuai dengan makna sebenarnya. Hal ini mengonfirmasi temuan pada tahap evaluasi bahwa pemahaman konteks yang lebih dalam masih menjadi tantangan bagi model.

IV. KESIMPULAN

Penelitian ini berhasil mengimplementasikan model IndoBERT untuk analisis sentimen ulasan wisata dengan akurasi pengujian mencapai 79%. Hasil evaluasi menunjukkan bahwa model sangat andal dalam mengidentifikasi sentimen positif (F1-score 0.87), namun menghadapi tantangan

pada sentimen netral dan negatif (F1-score 0.65). Kesenjangan performa antara data latih (akurasi 0.97) dan data uji (akurasi 0.79) mengindikasikan adanya *overfitting*. Selain itu implementasi juga berhasil dilakukan dengan mengintegrasikan model yang sudah di *fine-tuning* dengan *framework Streamlit* dengan tujuan agar model dapat digunakan oleh pihak yang berkepentingan.

Adapun beberapa tindakan yang bisa dilakukan untuk meningkatkan akurasi setiap kelas diantaranya adalah dengan melakukan *hyperparameter tuning* dengan bantuan pustaka Optuna, GridSearchCV, dan sejenisnya untuk membantu mencari parameter paling optimal untuk model atau dengan mengatur layer regulisasi seperti *dropout*, *learning rate* dan sebagainya akan sangat membantu untuk mengatasi *overfitting* yang dialami oleh model. Adapun tindakan lain seperti memperkaya dataset baik dengan cara augmentasi ataupun manual jika bisa diterapkan, selain itu melabeli ulang dataset juga bisa menjadi pertimbangan untuk meningkatkan nilai akurasi dari model secara keseluruhan.

Dengan menggunakan pendekatan *machine learning* terutama penggunaan *natural language processing* untuk prediksi sentimen ulasan wisata, terdapat banyak potensi dan manfaat yang bisa diambil. Hal ini membuka peluang bagi pemilik wisata ataupun pemangku kepentingan untuk memahami respon dari pengunjung secara dinamis dan akurat, dengan begitu pemangku kepentingan dapat mengambil keputusan yang tepat dan cepat untuk meningkatkan potensi wisata dan mampu bersaing dalam era teknologi.

Dengan demikian, pengembangan analisis sentimen yang diusulkan dalam penelitian ini dapat dijadikan referensi untuk pengembangan analisis sentimen lainnya yang menggunakan pendekatan *natural language processing* dan diintegrasikan dengan *framework* pendukung seperti *streamlit*.

UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua individu dan pihak yang telah memberikan kontribusi dan dukungan selama penelitian ini. Tanpa kerja

keras, panduan, dan bantuan mereka, penelitian ini tidak akan terwujud. Penulis ingin mengucapkan terima kasih kepada:

1. Pembimbing penulis, Ibu Hanifah Pertamasari, M.Kom, atas arahan, saran, dan pengawasannya yang berharga sepanjang penelitian ini.
2. Pembimbing penulis, Bapak Joni Maulindar, S.Kom., M.Eng, atas arahan, saran, dan pengawasannya yang berharga sepanjang penelitian ini.
3. Keluarga dan teman-teman penulis atas dukungan, motivasi, dan pengertian yang mereka berikan selama proses penelitian ini.
4. Institusi dan lembaga yang telah memberikan akses dan izin untuk mengumpulkan data yang diperlukan.

Terima kasih atas semua dukungan dan kontribusi yang telah diberikan. Semua bantuan ini sangat berarti bagi kelancaran dan kesuksesan penelitian penulis.

REFERENSI

- [1] B. Armutcu, A. Tan, M. Amponsah, S. Parida, dan H. Ramkissoon, "Tourist behaviour: The role of digital marketing and social media," *Acta Psychologica*, vol. 240, hlm. 104025, Okt 2023, doi: 10.1016/j.actpsy.2023.104025.
- [2] X. (Simon) Hu dan Y. Yang, "What makes online reviews helpful in tourism and hospitality? a bare-bones meta-analysis," *Journal of Hospitality Marketing & Management*, vol. 30, no. 2, hlm. 139–158, Feb 2021, doi: 10.1080/19368623.2020.1780178.
- [3] C. W. Park, I. Sutherland, dan S. K. Lee, "Effects of online reviews, trust, and picture-superiority on intention to purchase restaurant services," *Journal of Hospitality and Tourism Management*, vol. 47, hlm. 228–236, Jun 2021, doi: 10.1016/j.jhtm.2021.03.007.
- [4] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, dan M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," *Natural Language Processing Journal*, vol. 6, hlm. 100059, Mar 2024, doi: 10.1016/j.nlp.2024.100059.
- [5] R. Imamguluyev, "The Rise of GPT-3: Implications for Natural Language Processing and Beyond," *Int. J. Res. Publ. Rev.*, vol. 4, no. 3, hlm. 4893–4903, Mar 2023, doi: 10.55248/gengpi.2023.4.33987.
- [6] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes," *klik*, vol. 2, no. 4, hlm. 148–155, Feb 2022, doi: 10.30865/klik.v2i4.315.
- [7] F. A. Larasati, D. E. Ratnawati, dan B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," 2022.
- [8] E. Indrayuni, A. Nurhadi, dan D. A. Kristiyanti, "Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc," *FaktorExacta*, vol. 14, no. 2, hlm. 64, Agu 2021, doi: 10.30998/faktorexacta.v14i2.9697.
- [9] A. Karanikola, G. Davrazos, C. M. Liapis, dan S. Kotsiantis, "Financial sentiment analysis: Classic methods vs. deep learning models," *IDT*, vol. 17, no. 4, hlm. 893–915, Nov 2023, doi: 10.3233/IDT-230478.
- [10] M. Zhou, N. Duan, S. Liu, dan H.-Y. Shum, "Progress in Neural NLP: Modeling, Learning, and Reasoning," *Engineering*, vol. 6, no. 3, hlm. 275–290, Mar 2020, doi: 10.1016/j.eng.2019.12.014.
- [11] M. Wankhade, A. C. S. Rao, dan C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, hlm. 5731–5780, Okt 2022, doi: 10.1007/s10462-022-10144-1.
- [12] M. V. Koroteev, "BERT: A Review of Applications in Natural Language Processing and Understanding," 2021, *arXiv*. doi: 10.48550/ARXIV.2103.11943.
- [13] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," 2020, *arXiv*. doi: 10.48550/ARXIV.2011.00677.
- [14] L. Kent, C. Snider, J. Gopsill, dan B. Hicks, "Mixed reality in design prototyping: A systematic review," *Design Studies*, vol. 77, hlm. 101046, Nov 2021, doi: 10.1016/j.destud.2021.101046.
- [15] R. Wasik, "The Advantages & Disadvantages of Prototyping - Rapids Reproductions," The Advantages & Disadvantages of Prototyping. Diakses: 6 Juni 2025. [Daring]. Tersedia pada: <https://rapidsrepro.com/advantages-disadvantages-prototyping/>
- [16] Y. Akkem, B. S. Kumar, dan A. Varanasi, "Streamlit Application for Advanced Ensemble Learning Methods in Crop Recommendation Systems – A Review and Implementation," *IJST*, vol. 16, no. 48, hlm. 4688–4702, Des 2023, doi: 10.17485/IJST/v16i48.2850.
- [17] A. Metonosa Afinda, "Lexicon Dataset." <https://github.com/angelmetanosaa/dataset>, Github, 12 Juni 2025. Diakses: 12 Juni 2025. [Daring]. Tersedia pada: <https://github.com/angelmetanosaa/dataset>