

Implementasi Bidirectional Encoder Representations from Transformers (BERT) untuk Analisis Sentimen Ulasan Aplikasi Ibis Paint X

Esti Setiyaningsih¹, Ety Sutanty², Maukar^{3*}

¹Program Studi Informatika, Universitas Gunadarma, Depok, Indonesia

²Program Studi Sistem Informasi, Universitas Gunadarma, Depok, Indonesia

³Program Studi Magister Manajemen Sistem Informasi, Universitas Gunadarma, Depok, Indonesia

Artikel Info

Kata Kunci:

Analisis Sentimen;
BERT;
Ibis Paint X;
Pelabelan;
Transformers.

Keywords:

Sentiment Analysis;
BERT;
Ibis Paint X;
Labelling;
Transformers.

Riwayat Artikel:

Submitted: 22 September 2025
Accepted: 18 November 2025
Published: 24 November 2025

Abstrak: Perkembangan aplikasi seni digital memunculkan banyak ulasan pengguna yang bermanfaat bagi pengembang, namun sulit dievaluasi secara manual. Penelitian ini menganalisis sentimen 8.500 ulasan berbahasa Indonesia pada aplikasi Ibis Paint X yang diperoleh melalui proses scraping Google Play Store menggunakan google-play-scraper. Dataset terdiri atas dua kolom utama, yaitu *review_text* dan *category* (rating), kemudian melalui tahap preprocessing mencakup pelabelan, pembersihan teks, penghapusan stopwords, tokenisasi, serta normalisasi. Setelah dibagi menjadi data latih (70%), validasi (15%), dan uji (15%), model IndoBERT di-*fine-tuning* selama lima epoch untuk melakukan klasifikasi sentimen positif, negatif, dan netral. Evaluasi performa model dilakukan menggunakan akurasi, precision, recall, F1-score, serta confusion matrix. Hasil pengujian menunjukkan akurasi sebesar 87%. Kinerja model paling tinggi pada kelas sentimen negatif dengan precision 0,94, recall 0,96, dan F1-score 0,95. Namun, performa pada kelas positif (F1-score 0,52) dan netral (F1-score 0,31) masih rendah, sebagaimana turut tercermin pada confusion matrix yang menunjukkan banyaknya kesalahan klasifikasi menuju kelas positif yang dominan. Temuan ini menegaskan efektivitas BERT dalam analisis sentimen berbahasa Indonesia, sekaligus menunjukkan perlunya penanganan ketidakseimbangan data untuk meningkatkan performa pada seluruh kategori sentimen.

Abstract: The growth of digital art applications has generated numerous user reviews that are valuable for developers but difficult to evaluate manually. This study analyzes the sentiment of 8,500 Indonesian-language reviews of the Ibis Paint X application, obtained through Google Play Store scraping using the google-play-scraper library. The dataset consists of two main columns, *review_text* and *category* (rating), and undergoes preprocessing steps including labeling, text cleaning, stopword removal, tokenization, and normalization. After being divided into training (70%), validation (15%), and test (15%) sets, the IndoBERT model is fine-tuned for five epochs to classify reviews into positive, negative, and neutral sentiments. Model performance is evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The test results show an accuracy of 87%. The model achieves its best performance on the negative sentiment class, with a precision of 0.94, recall of 0.96, and F1-score of 0.95. However, performance on the positive (F1-score 0.52) and neutral (F1-score 0.31) classes remains low, which is also reflected in the confusion matrix showing frequent misclassification toward the dominant positive class. These findings highlight the effectiveness of BERT for Indonesian-language

sentiment analysis while emphasizing the need to address data imbalance to improve performance across all sentiment categories.

Corresponding Author:

Maukar

Email: maukar@staff.gunadarma.ac.id

PENDAHULUAN

Perkembangan teknologi informasi telah memberikan dampak signifikan dalam berbagai bidang, termasuk seni dan kreativitas (Li et al., 2024). Salah satu wujud perkembangan tersebut adalah hadirnya aplikasi mobile berbasis seni yang memungkinkan pengguna mengekspresikan diri melalui media digital (Wang & Zhang, 2024). Ibis Paint X merupakan salah satu aplikasi populer yang hingga Juli 2024 telah diunduh lebih dari 370 juta kali di Google Play Store, menunjukkan tingginya tingkat adopsi dan minat masyarakat global terhadap platform ini (Hasanudin et al., 2021). Tingginya jumlah unduhan diikuti dengan meningkatnya ulasan pengguna yang mencerminkan pengalaman, kepuasan, maupun keluhan terhadap aplikasi. Ulasan tersebut tidak hanya menjadi sarana komunikasi antara pengguna dan pengembang, tetapi juga berfungsi sebagai sumber data penting yang dapat memberikan gambaran objektif mengenai kualitas aplikasi.

Tantangan yang muncul adalah besarnya volume ulasan yang menyebabkan proses evaluasi manual menjadi tidak efisien. Pengembang aplikasi sering kali kesulitan mengidentifikasi pola kepuasan maupun permasalahan utama secara cepat dan akurat (Nawawi et al., 2024). Hal ini menjadikan analisis sentimen berbasis kecerdasan buatan, khususnya model Bidirectional Encoder Representations from Transformers (BERT), relevan untuk memberikan pemetaan sentimen yang lebih terstruktur (Mustak et al., 2024). Analisis sentimen terhadap ulasan pengguna Ibis Paint X menjadi sangat penting untuk memahami persepsi pengguna secara menyeluruh dan mendukung pengembangan aplikasi yang lebih responsif.

Penelitian terkait implementasi model BERT dalam analisis sentimen telah dilakukan oleh beberapa peneliti sebelumnya. Penelitian (Alaparthi & Mishra, 2020) melakukan evaluasi komparatif terhadap pendekatan berbasis leksikon seperti Sent WordNet, metode machine learning tradisional seperti logistic regression, model LSTM, dan model lanjutan BERT menggunakan dataset publik IMDB berjumlah 50.000 ulasan. Penelitian ini menunjukkan bahwa BERT consistently outperforms model lain. Penelitian (Pradipta & Widodo, 2024) mengimplementasikan BERT untuk analisis sentimen komentar masyarakat di Instagram dan Facebook terkait pandemi Covid-19 varian Omicron, dengan akurasi terbaik mencapai 90%. Penelitian (Maulana et al., 2025) mengaplikasikan IndoBERT untuk analisis sentimen publik terhadap program naturalisasi pemain Tim Nasional Indonesia melalui data Twitter, menunjukkan akurasi 90% dengan hasil metrik yang kompetitif. Sementara itu, penelitian (Areshey & Mathkour, 2023) membandingkan model BERT yang telah di-*fine-tuning* dengan KNN, Naïve Bayes, dan SVM untuk klasifikasi ulasan pelanggan dan secara jelas memperlihatkan superioritas BERT dalam akurasi dan stabilitas.

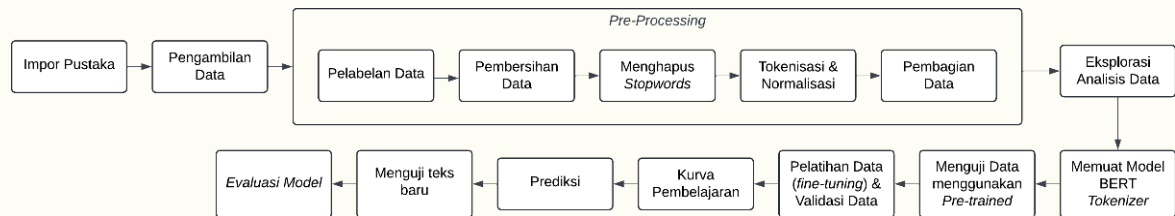
Meskipun kajian pustaka tersebut menunjukkan kemampuan BERT dalam memahami konteks dan melakukan klasifikasi sentimen, identifikasi *research gap* tetap muncul. Berbagai penelitian menggunakan domain data yang berbeda dari aplikasi seni digital, seperti ulasan film (Alaparthi & Mishra, 2020), komentar pandemi (Pradipta & Widodo, 2024), atau sentimen terhadap kebijakan olahraga (Maulana et al., 2025), sehingga belum merepresentasikan karakteristik bahasa dan ekspresi pengguna aplikasi kreatif. Mayoritas penelitian memanfaatkan data berskala kecil atau terbatas platform seperti Instagram, Facebook, dan Twitter (Pradipta & Widodo, 2024; Maulana et al., 2025), padahal ulasan aplikasi pada Google Play Store memiliki pola bahasa, konteks, dan kebutuhan analisis yang berbeda. Sebagian studi seperti (Areshey & Mathkour, 2023) masih berfokus pada perbandingan algoritma tanpa menerapkan BERT pada domain yang lebih spesifik. Kajian sebelumnya cenderung

deskriptif dan belum membahas bagaimana BERT bekerja pada konteks aplikasi seni digital yang memiliki karakteristik unik seperti bahasa evaluatif, teknis, dan ekspresif.

Dengan mempertimbangkan celah tersebut, penelitian ini memposisikan diri sebagai studi yang menerapkan model BERT (Oliaee et al., 2023) pada domain ulasan aplikasi seni digital, yaitu Ibis Paint X, yang hingga kini belum banyak dieksplorasi. Penelitian ini memanfaatkan 8.500 ulasan dari Google Play Store, menerapkan preprocessing komprehensif, serta melakukan evaluasi melalui akurasi, presisi, recall, dan F1-score. Dengan demikian, penelitian ini tidak hanya melengkapi kekurangan studi sebelumnya, tetapi juga memperluas cakupan aplikasi BERT pada domain kreatif digital, serta memberikan wawasan empiris mengenai persepsi pengguna aplikasi seni.

METODE

Penelitian terdiri dari beberapa tahapan dalam analisis sentimen aplikasi Ibis Paint X dengan mengimplementasikan model BERT seperti dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian Implementasi Model BERT pada Ibis Paint X

Gambar 1 menampilkan alur tahapan penelitian analisis sentimen ulasan pada aplikasi Ibis Paint X. Tahapan pertama dimulai dengan impor pustaka yang diperlukan untuk pemrosesan data dan model. Setelah itu, data ulasan aplikasi Ibis Paint X dikumpulkan dari Google Play Store. Data yang telah terkumpul kemudian melalui tahap *preprocessing* yang mencakup pelabelan manual, pembersihan dari karakter berulang, emoji, link, simbol, dan spasi, penghapusan *stopwords*, serta tokenisasi dan normalisasi teks. Data yang sudah bersih kemudian dibagi menjadi data pelatihan, data validasi, dan data uji dengan data pelatihan (train set) 70%, data validasi (validation set) 15%, dan data pengujian (test set) 15%. Tahap selanjutnya adalah eksplorasi dan analisis data awal, visualisasi pie chart digunakan untuk menggambarkan proporsi kategori sentimen. Setelah itu, model BERT tokenizer dimuat dan data diuji menggunakan model pre-trained. Tahap pelatihan dilakukan *fine-tuning* pada model menggunakan data pelatihan, dan hasilnya divalidasi menggunakan data validasi. Kurva pembelajaran dianalisis untuk memantau perkembangan kinerja model selama pelatihan. Setelah model selesai dilatih, dilakukan prediksi terhadap data baru, diikuti dengan evaluasi model menggunakan *confusion matrix* dan laporan klasifikasi untuk mengukur akurasi dan efektivitas prediksi model.

Impor Pustaka

Tahap awal penelitian adalah mengimpor pustaka Python yang mendukung seluruh proses analisis sentimen, mulai dari pengambilan data, preprocessing, pelatihan model, hingga visualisasi hasil. Pustaka yang digunakan antara lain:

1. Pengambilan Data: `google_play_scraper` untuk mengekstrak ulasan aplikasi dari Google Play Store (Latif et al., 2019).
2. *Preprocessing* Teks: `regex`, `re`, `Emoji`, `Sastrawi`, `nlk` (tokenisasi, *stopwords*), serta modul `sent_tokenize` dan `word_tokenize` (Albab et al., 2023).
3. Manipulasi Data: `pandas`, `numpy`, dan `defaultdict` dari `collections` (Rayhan et al., 2023) untuk pengolahan dataset.
4. Visualisasi: `matplotlib`, `seaborn`, dan `WordCloud` untuk menampilkan distribusi data dan representasi kata.
5. Pemodelan NLP: `transformers` (Hugging Face), `torch`, dan `IndoBERT` untuk implementasi model

6. Modul Tambahan: scikit-learn untuk pembagian data dan evaluasi model, tqdm untuk *progress bar*, serta modul indonlu (berserta utilitasnya seperti *DocumentSentimentDataset*, *DataLoader*, dan *metrics_fn*) (Cahyawijaya et al., 2021) untuk mendukung pemrosesan bahasa alami berbahasa Indonesia.

1. *Scraping* Data. Pada penelitian ini menggunakan pustaka google-play-scraper. Tentukan parameter: `app_id = 'jp.ne.ibis.ibispaintx.app'`, `lang='id'` (bahasa Indonesia), `country='id'` (Indonesia), `sort=Sort.MOST_RELEVANT` (relevansi tertinggi), `count=8500` (jumlah ulasan).
2. Menyimpan Data. Data ulasan (teks dan rating) hasil *scraping* disimpan ke file .csv.
3. Memuat ke DataFrame (pandas) dilakukan dengan : (1) Baca kembali file .csv ke dalam DataFrame. (2) Ubah nama kolom: *content* menjadi *review_text* (isi ulasan) dan *score* menjadi *category* (rating).
4. Hasil akhir dataset terdiri dari 8500 baris dengan 2 kolom utama: *review_text* merupakan teks ulasan pengguna dan *category* merupakan skor/rating pengguna.

Data hasil scraping memiliki skor rating dari 1 hingga 5, yang kemudian diubah menjadi kategori sentimen. Skor 1 dan 2 dikategorikan sebagai sentimen negatif, skor 3 sebagai netral, sedangkan skor 4 dan 5 sebagai positif. Proses pelabelan ini menghasilkan distribusi data yang cukup timpang, dengan mayoritas ulasan berlabel positif sebanyak 7373 data, sementara label negatif berjumlah 595 data, dan label netral 532 data. Informasi ini penting karena distribusi kategori memengaruhi performa model saat dilatih. Pelabelan data dalam bentuk .csv dapat dilihat pada Gambar 2.

[illegible]

2. Pembersihan data (*data cleaning*).

Pada tahap ini, teks ulasan dibersihkan dari elemen-elemen yang tidak relevan seperti huruf kapital berlebihan, emoji, emotikon, tautan, username, hashtag, angka, simbol, serta karakter yang berulang (Hadiprakoso et al., 2023). Semua teks juga diubah ke dalam huruf kecil (*case folding*) untuk menjaga konsistensi. Melalui pembersihan ini, teks ulasan menjadi lebih sederhana dan siap diproses lebih lanjut. Tabel 1 menampilkan perbandingan antara teks ulasan sebelum dan sesudah proses pembersihan data.

Tabel 1. Perbandingan Sebelum dan Sesudah Pembersihan Data

Sebelum Data Cleaning	Setelah Data Cleaning
⚡ Seru Banget ⚡ 1 Gambarnya Bagus 2 Wow 3 Gambar 🙌 4 Sos 5 Lama 6 Tari 7 Indah 8 Ilkan 9 Jika Main Seru 10 Mejik	seru banget gambarnya bagus wow gambar sos lama tari indah ilkan jika main seru mejik
Saan sedavn meng instal ini selau terulang ulang padahal sudah 83% hampir selesai instal tapi selalu saja suka mengulang ulang bikin kesel 🤬 sebelllllll.....iiiihhhhhhh....	saan sedavn meng instal ini selau terulang ulang padahal sudah hampir selesai instal tapi selalu saja suka mengulang ulang bikin kesel sebell iihh

3. Penghapusan stopwords.

Stopwords adalah kata-kata umum yang tidak memiliki bobot signifikan dalam analisis sentimen, seperti “dan”, “yang”, atau “dengan”. Daftar *stopwords* yang digunakan berasal dari *stopwords_id_satya* yang diunduh dari GitHub (Diaz, 2025), kemudian ditambahkan dengan stopwords tambahan sesuai kebutuhan penelitian. Proses menghapus stopwords ini, membuat teks ulasan menjadi lebih ringkas dan hanya berisi kata-kata yang lebih bermakna untuk analisis. Tabel 2 menunjukkan perbandingan sebelum dan sesudah menghapus *stopwords*.

Tabel 2. Perbandingan Sebelum dan Sesudah Menghapus Stopwords

Sebelum Remove Stopwords	Setelah Remove Stopwords
seru banget gambarnya bagus wow gambar sos lama tari indah ilkan jika main seru mejik	seru banget gambarnya bagus wow gambar sos tari indah ilkan main seru mejik
saan sedavn meng instal ini selau terulang ulang padahal sudah hampir selesai instal tapi selalu saja suka mengulang ulang bikin kesel sebell iihh	saan sedavn meng instal selau terulang ulang selesai instal suka mengulang ulang bikin kesel sebell iihh

4. Tokenisasi dan Normalisasi.

Tokenisasi dilakukan dengan memecah teks ulasan menjadi potongan-potongan kata (token). Setelah teks dipecah, dilakukan normalisasi untuk mengganti kata-kata tidak baku atau bahasa gaul menjadi bentuk formal sesuai kamus normalisasi (kamus alay) (Aliero et al., 2023). Contohnya, kata “install” yang salah ketik menjadi “instal” akan disesuaikan agar konsisten dengan bahasa formal. Proses ini membantu menyatukan berbagai variasi penulisan sehingga data lebih seragam.

	Setelah Tokenize	Setelah Normalize
seru banget gambarnya bagus wow gambar sos lama tari indah ilkan jika main seru mejik	['seru', 'banget', 'gambarnya', 'bagus', 'wow', 'gambar', 'sos', 'lama', 'tari', 'indah', 'ilkan', 'jika', 'main', 'seru', 'mejik']	seru banget gambarnya bagus wow gambar sos lama tari indah ilkan jika main seru mejik

saan sedanv meng instal ini selau terulang ulang padahal sudah hampir selesai instal tapi selalu saja suka mengulang ulang bikin kesel sebell iihh	['saan', 'sedanv', 'meng', 'instal', 'ini', 'selau', 'terulang', 'ulang', 'padahal', 'sudah', 'hampir', 'selesai', 'instal', 'tapi', 'selalu', 'saja', 'suka', 'mengulang', 'ulang', 'bikin', 'kesel', 'sebell', 'iihh']	saan sedanv meng install ini selau terulang ulang padahal sudah hampir selesai install tapi selalu saja suka mengulang ulang bikin kesel sebell iihh
---	---	---

Tabel 3. Perbandingan Sebelum dan Sesudah Tokenisasi dan Normalisasi

Tabel 3 menampilkan perbandingan teks ulasan sebelum dan sesudah proses tokenisasi dan normalisasi. Tahap tokenisasi, teks ulasan dipecah menjadi kata-kata atau token individual. Selanjutnya, dilakukan normalisasi teks, yaitu proses mengganti kata-kata tidak baku atau informal dengan kata-kata yang lebih formal. Proses ini menggunakan kamus alay yang dimuat dari file *kamus_alay.csv*, yang diambil dari GitHub. Normalisasi ini penting untuk menyatukan berbagai variasi penulisan yang sering ditemukan dalam bahasa sehari-hari atau slang.

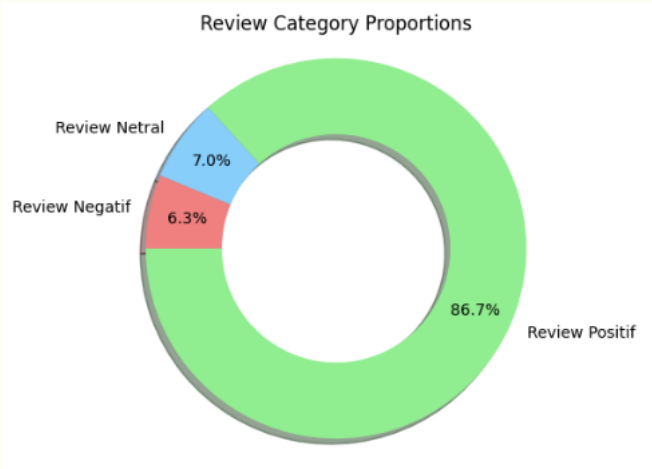
5. Pembagian data (*splitting data*).

Dataset yang telah diproses dibagi menjadi tiga bagian: data pelatihan (70%), data validasi (15%), dan data pengujian (15%). Proses pembagian dilakukan dengan metode *stratified* agar proporsi setiap kategori sentimen tetap seimbang di masing-masing subset. Pembagian ini penting untuk memastikan model tidak hanya belajar dari satu kategori dominan, melainkan dapat diuji secara seimbang dengan data yang beragam.

Melalui rangkaian preprocessing ini, data ulasan aplikasi Ibis Paint X berhasil diubah menjadi dataset yang bersih, terstruktur, dan siap digunakan dalam tahap analisis sentimen. Dataset akhir terdiri dari teks ulasan yang telah dinormalisasi pada kolom *review_text* dan label sentimen pada kolom *category*, kemudian disimpan dalam file TSV agar mudah dipanggil kembali untuk tahap pelatihan model.

Eksplorasi Analisis Data

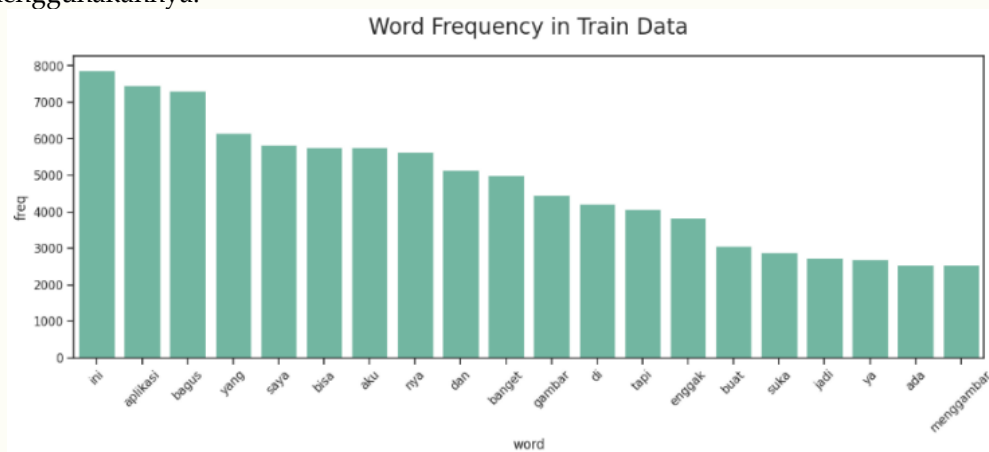
Tahap eksplorasi analisis data dilakukan untuk memahami struktur dan karakteristik ulasan sebelum masuk ke tahap pemodelan. Langkah pertama adalah melihat distribusi kategori sentimen yang diperoleh dari proses pelabelan. Dari total 8500 data, sebanyak 7373 ulasan termasuk dalam kategori positif (86,72%), 595 ulasan negatif (7,01%), dan 532 ulasan netral (6,25%). Distribusi ini divisualisasikan dengan *Donut chart* Gambar 3, yang memperlihatkan dominasi ulasan positif dalam dataset. Visualisasi ini membantu menunjukkan bahwa mayoritas pengguna aplikasi Ibis Paint X memberikan tanggapan yang baik terhadap aplikasi.



Gambar 3. Donut Chart Karakteristik Ulasan

Setelah itu, dilakukan analisis lebih dalam terhadap teks ulasan dengan membentuk sebuah

corpus, yaitu gabungan seluruh teks review yang telah melalui tahap *preprocessing*. Dari corpus ini dihitung frekuensi kata, sehingga dapat diketahui kata-kata yang paling sering muncul. Hasilnya divisualisasikan dalam bentuk grafik batang Gambar 4 yang menampilkan 20 kata teratas. Kata seperti *aplikasi*, *bagus*, *banget*, *saya*, dan *gambar* muncul dengan frekuensi tinggi, mencerminkan topik utama yang dibicarakan oleh pengguna, yaitu fungsi aplikasi, kualitas gambar, serta pengalaman pribadi dalam menggunakannya.



Gambar 4. Tampilan Bagan Bar Kata yang Paling Sering Muncul

Melalui eksplorasi analisis data ini, diperoleh gambaran umum mengenai persebaran sentimen dalam dataset dan kata-kata yang paling sering digunakan oleh pengguna. Informasi ini bermanfaat untuk memahami konteks data sekaligus sebagai dasar pertimbangan dalam tahap pemodelan analisis sentimen selanjutnya.

Implementasi Model BERT Tokenizer

Setelah model tokenizer dimuat, dilakukan penyesuaian konfigurasi model IndoBERT untuk tugas klasifikasi sentimen. Konfigurasi model yang digunakan dalam penelitian ini ditampilkan pada Tabel 4. Tabel ini merangkum parameter pelatihan utama, termasuk arsitektur model, jumlah epoch, ukuran batch, panjang maksimum sequence, serta optimizer dan metrik evaluasi yang digunakan.

Tabel 4. Konfigurasi Model BERT untuk Analisis Sentimen Ibis Paint X

Parameter	Nilai
Model dasar	indobenchmark/indobert-base-p1
Arsitektur	BertForSequenceClassification
Jumlah label	3 (positif, negatif, netral)
Optimizer	AdamW
Learning rate	2e-5
Batch size	32
Max sequence length	512
Epoch	5
Warmup steps	0
Dropout hidden layers	0.1
Padding	True (dynamic)
Truncation	True
Tokenizer	IndoBERT tokenizer
Loss function	CrossEntropyLoss
Metric evaluasi	Accuracy, Precision, Recall, F1-score, Confusion Matrix

Tahap ini bertujuan untuk memuat dan menginisialisasi BERT Tokenizer dari model pralatih indobenchmark/indobert-base-p1, yang akan digunakan untuk analisis sentimen. Tokenizer ini

berfungsi mengubah teks ulasan mentah menjadi format numerik yang sesuai untuk dimasukkan ke dalam model BERT. Proses tokenisasi mencakup pemisahan teks menjadi token, penambahan token khusus seperti [CLS] di awal dan [SEP] di akhir teks, serta penerapan *padding* dan *truncation* agar panjang input konsisten sesuai batas maksimum (*max_seq_len*=512). Selain itu, *tokenizer* juga menghasilkan *attention mask* yang membantu model membedakan antara token asli dan *padding*.

Setelah *tokenizer* siap, langkah selanjutnya adalah memuat konfigurasi model BERT yang telah dilatih sebelumnya. Model dasar *indobert-base-p1* kemudian diadaptasi menjadi *BertForSequenceClassification* dengan jumlah label disesuaikan dengan kategori sentimen pada dataset (positif, negatif, dan netral). Dataset yang telah diproses sebelumnya diorganisasikan ke dalam tiga bagian: train, validation, dan test set. Masing-masing dataset ini dimuat menggunakan *DocumentSentimentDataset* dan *DocumentSentimentDataLoader* dengan parameter tertentu, misalnya *batch_size*=32 untuk jumlah data per *batch*, serta *shuffle*=True pada data latih agar distribusi data lebih acak selama proses *training*. Melalui tahap ini, teks ulasan yang semula tidak terstruktur berhasil dipetakan menjadi representasi numerik yang dapat dipahami oleh model BERT. Dengan demikian, model siap digunakan dalam proses pelatihan dan evaluasi analisis sentimen pada data ulasan aplikasi Ibis Paint X.

Menguji Data dengan Model Pre-trained

Tahap ini dilakukan dengan menggunakan model BERT pre-trained yang telah dimuat sebelumnya. Data uji yang belum pernah dilihat oleh model diproses melalui *tokenizer* untuk diubah menjadi token numerik, kemudian dimasukkan ke dalam model. Model menghasilkan *logits*, yaitu skor mentah untuk setiap kelas sentimen, dan label prediksi ditentukan berdasarkan skor tertinggi. Tahap ini penting untuk mengevaluasi kemampuan awal model dalam mengenali pola sentimen tanpa adanya proses pelatihan tambahan (*fine-tuning*).

Pelatihan (Fine-tuning) dan Validasi Model

Agar model dapat menyesuaikan diri dengan dataset ulasan Ibis Paint X, dilakukan proses *fine-tuning* menggunakan optimizer AdamW dengan *learning rate* kecil agar pembaruan parameter berlangsung stabil. Proses pelatihan dilakukan dalam beberapa *epoch*, di mana pada setiap *epoch* model memproses data latih, menghitung *loss*, melakukan *backpropagation*, dan memperbarui bobot. Selain pelatihan, model juga dievaluasi pada data validasi setiap *epoch*. Validasi dilakukan tanpa pembaruan bobot untuk memastikan model tidak mengalami *overfitting* dan tetap mampu melakukan generalisasi dengan baik. Metrik evaluasi seperti akurasi, precision, recall, dan f1-score dicatat untuk memantau performa model. Untuk memvisualisasikan proses pelatihan, dibuat kurva pembelajaran yang menampilkan perubahan akurasi pada data latih dan validasi sepanjang *epoch*. Grafik ini membantu mengidentifikasi apakah model mengalami *underfitting* (jika akurasi rendah di kedua dataset) atau *overfitting* (jika akurasi tinggi di latih tapi rendah di validasi).

Setelah model selesai dilatih, tahap berikutnya adalah prediksi pada data uji maupun teks baru. Model dijalankan dalam mode evaluasi (*model.eval()*) agar hasilnya konsisten. Untuk teks baru, data terlebih dahulu di-tokenisasi, kemudian diproses oleh model untuk menghasilkan label sentimen. Selain label, probabilitas untuk setiap kelas dihitung menggunakan fungsi *softmax* sehingga tingkat keyakinan model terhadap prediksi dapat diketahui. Hal ini memungkinkan interpretasi lebih jelas terhadap hasil klasifikasi.

HASIL DAN PEMBAHASAN

Pada penelitian ini, *preprocessing* dilakukan untuk memastikan bahwa data ulasan siap digunakan dalam pelatihan model analisis sentimen. Tahapan ini meliputi pelabelan data, pembersihan teks, case folding, tokenisasi, normalisasi, hingga pembagian data ke dalam subset pelatihan, validasi, dan pengujian. Tabel 5 menampilkan contoh hasil *preprocessing*. Dari tabel tersebut terlihat bahwa teks ulasan yang awalnya masih mengandung berbagai simbol, emotikon, hingga karakter khusus berhasil dibersihkan sehingga menjadi teks yang lebih sederhana dan konsisten. Misalnya, emotikon “(‘3`·)”

serta simbol hati “~♥~” pada ulasan pertama berhasil dihapus dan diganti dengan teks bersih. Hal ini menunjukkan bahwa proses *cleaning* cukup efektif dalam mengurangi noise pada data.

Tabel 5. Hasil *Preprocessing* Data

Sebelum <i>Preprocessing</i>	Setelah <i>Preprocessing</i>	Sentimen
Aplikasinya membantu banget (·'3`·), bagi kita yg suka menggambar. Pilihan kuas yang bervariasi juga sangat membantu. Fitur ² yang sangat membantu juga pastinya. Ada benerapa bug saat masuk ke dalam aplikasi, mungkin bisa di perbaiki lagi.. Untuk sisanya.. oke semua dan gak ada keluhan lain... ~♥~	aplikasinya membantu banget bagi kita yang suka menggambar pilihan kuas yang bervariasi juga sangat membantu fitur yang sangat membantu juga pastinya ada benerapa bug saat masuk ke dalam aplikasi mungkin bisa di perbaiki lagi untuk sisanya oke semua dan enggak ada keluhan lain	Positif
Aplikasinya bagus banget>_< gambarku jadi bisa ku edit+lebh keren (J°□°)tapi harus diperbaiki lagi ya karena punya suka ngelag pas lagi ngambar/buat kanvas baru jadi tolong diperbaiki 🙏🙏🙏😊😊	aplikasinya bagus banget gambarku jadi bisa ku edit lebh keren tapi harus diperbaiki lagi ya karena punya suka ngelag pas lagi ngambar buat kanvas baru jadi tolong diperbaiki	Positif
Apa aku aja yang ngerasa makin kesini makin kaku kalau nge rotate kanvas nya yah?? Padahal itu menurut ku penting bagi para artist buat cek art nya dan ngemudahin buat sketch atau ngewarnain, kalau sekarang susah aja ngerotate nya, kaku banget	apa aku saja yang merasa makin kesini makin kaku kalau nge rotate kanvas nya ya padahal itu menurut ku penting bagi para artist buat cek art nya dan ngemudahin buat sketch atau ngewarnain kalau sekarang susah saja ngerotate nya kaku banget	Netral
Untuk yang lain ini bagus cuman di bagian animasi saat barkas ulang malah keluar aplikasi padahal sudah mau 100% tapi malah keluar aplikasi otomatis	untuk yang lain ini bagus cuman di bagian animasi saat barkas ulang malah keluar aplikasi padahal sudah mau tapi malah keluar aplikasi otomatis	Negatif
Update nya gak bagus banget! Setiap mau coret pasti nambah lapisan vector nambah nambah layer bisa bisa kebanyakan layer karena update ini jelek banget ini game vektor vektor terus!!! 🙏🙏🙏	update nya enggak bagus banget setiap mau coret pasti menambah lapisan vector menambah layer bisa bisa kebanyakan layer karena update ini jelek banget ini game vektor vektor terus	Negatif

Seperti dapat dilihat pada Tabel 5, masih terdapat beberapa kata yang tidak sepenuhnya tersaring, seperti kata umum “ini” yang tetap muncul meskipun dianggap *stopword*. Hal ini menandakan bahwa daftar *stopwords* dan kaus alay yang digunakan masih dapat diperluas atau disesuaikan lebih lanjut agar hasil *preprocessing* semakin optimal. Kendati demikian, secara keseluruhan teks hasil *preprocessing* sudah lebih siap digunakan untuk analisis karena strukturnya lebih rapi, konsisten, dan bebas dari simbol yang tidak relevan. Selanjutnya, data yang telah diproses dibagi menjadi tiga subset untuk keperluan pelatihan dan evaluasi model. Dari total 8000 data mentah, sebanyak 5941 data (74,26%) digunakan sebagai data latih, 1274 data (15,93%) digunakan sebagai data validasi, dan 1273 data (15,91%) digunakan sebagai data uji. Pembagian ini penting karena menjamin bahwa model tidak hanya belajar dari data latih, tetapi juga diuji pada data baru yang belum pernah dilihat sebelumnya. Melalui proporsi ini, model diharapkan dapat memperoleh kemampuan generalisasi yang baik serta mengurangi risiko *overfitting*.

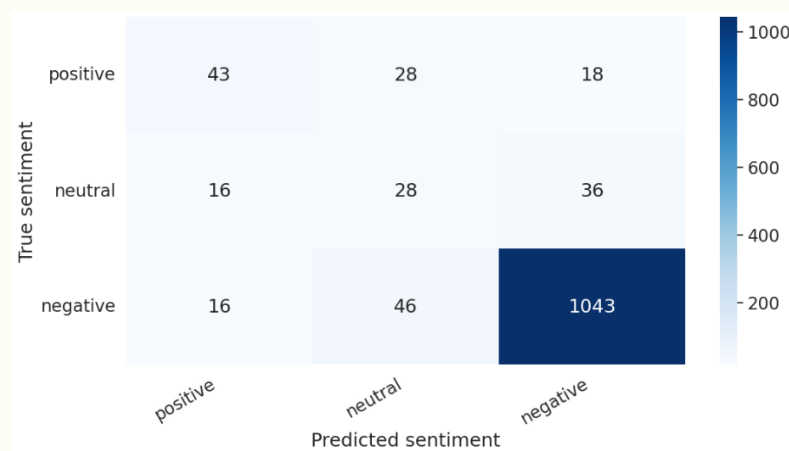
Hasil Pelatihan Model BERT

Proses pelatihan dilakukan selama 5 epoch dengan durasi total sekitar 90 menit 63 detik, rata-rata 18 menit per epoch. Hasil pelatihan ditunjukkan pada Tabel 6, yang memuat data akurasi dan loss untuk pelatihan dan validasi. Terlihat bahwa *train accuracy* meningkat konsisten dari 0,88 pada epoch pertama hingga mencapai 0,97 pada epoch terakhir. Hal ini menandakan bahwa model BERT semakin mampu mengenali pola pada data pelatihan. Sebaliknya, *validation accuracy* cenderung fluktuatif dan bahkan menurun pada epoch akhir (0,88). Tren ini diperkuat dengan peningkatan *validation loss* dari 0,2778 menjadi 0,4107, yang menunjukkan indikasi *overfitting*, dimana model terlalu menyesuaikan diri dengan data latih, sehingga kurang mampu menggeneralisasi pada data validasi.

Tabel 6. Tabel Data Akurasi dan Loss

Epoch	Train Acc	Val Acc	Train Loss	Val Loss
1	0.88	0.90	0.3268	0.2778
2	0.91	0.89	0.2425	0.2655
3	0.93	0.90	0.1924	0.2710
4	0.95	0.89	0.1446	0.3344
5	0.97	0.88	0.0884	0.4107

Nilai F1-score naik dari 0,48 menjadi 0,89, menunjukkan bahwa model semakin baik dalam menyeimbangkan presisi dan recall pada data latih. Berbeda dengan training, hasil validasi relatif stagnan, dengan F1-score hanya berkisar antara 0,54–0,60. Setelah proses pelatihan dan validasi model selesai, evaluasi model dilakukan untuk mengukur kinerjanya dalam analisis sentimen ulasan aplikasi menggunakan BERT. Evaluasi awal dilakukan dengan membuat *confusion matrix* untuk melihat distribusi prediksi model terhadap tiga kategori sentimen: positif, netral, dan negatif. *Confusion matrix* memudahkan identifikasi jumlah prediksi yang benar maupun salah, sekaligus memberikan gambaran area yang masih perlu perbaikan. Gambar 5 menampilkan *confusion matrix* hasil pengujian pada data uji. Model cukup andal dalam mengenali ulasan negatif dengan *True Negative* yang tinggi (1043) dan *False Negative* relatif rendah (54). Namun, performa melemah pada kategori netral, hanya menghasilkan 28 *True Neutral* dibandingkan 74 *False Neutral*. Selain itu, terdapat 32 *False Positive*, yang menunjukkan model masih sering mengelompokkan ulasan positif atau netral sebagai negatif. Secara keseluruhan, model efektif dalam mendeteksi sentimen negatif, tetapi kurang akurat dalam membedakan ulasan positif dan netral.

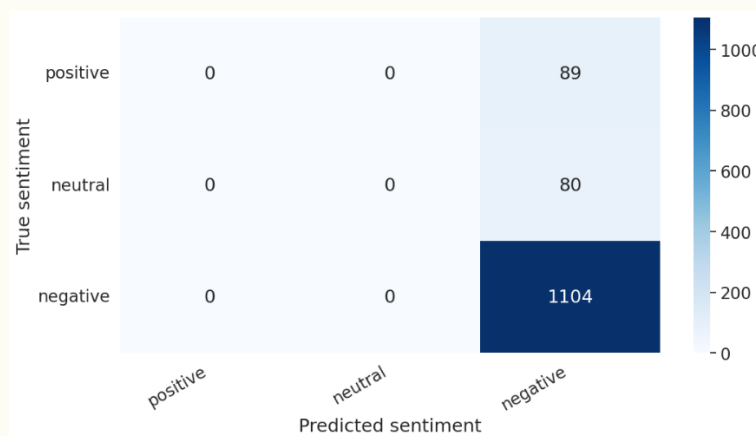
Gambar 5. *Confusion Matrix Testing Data*

Selain confusion matrix, evaluasi kinerja model pada data uji juga dihitung menggunakan metrik precision, recall, dan F1-score untuk masing-masing kelas sentimen. Hasil lengkap evaluasi ditampilkan pada Tabel 7.

Tabel 7. Evaluasi Kinerja Model BERT pada Data Uji

Kelas Sentimen	Precision	Recall	F1-Score	Support
Positif	0.63	0.44	0.52	7373
Netral	0.40	0.23	0.31	532
Negatif	0.94	0.96	0.95	595
Rata-rata Makro	0.66	0.54	0.59	–
Akurasi	–	–	0.87	8.500

Gambar 6 memperlihatkan hasil pada data validasi. Model menunjukkan bias yang sangat kuat terhadap kategori negatif, dengan *True Negative* sebesar 1104, sementara *True Positive* dan *True Neutral* bernilai nol. Kondisi ini mengakibatkan *False Negative* cukup tinggi (169), menandakan bahwa model salah mengklasifikasikan seluruh ulasan positif dan netral sebagai negatif. Hasil ini menunjukkan keterbatasan serius dalam kemampuan generalisasi model, sehingga diperlukan perbaikan, misalnya melalui penyeimbangan data atau strategi *fine-tuning* tambahan.

Gambar 6. *Confusion Matrix Validasi Data*

KESIMPULAN

Penelitian ini menunjukkan bahwa model BERT, khususnya IndoBERT, mampu digunakan secara efektif untuk analisis sentimen ulasan aplikasi seni digital berbahasa Indonesia. Dengan dataset 8.500 ulasan pengguna Ibis Paint X, model yang di-fine-tuning selama lima epoch berhasil mencapai akurasi 87% pada data uji. Hasil ini memperlihatkan bahwa BERT memiliki keunggulan dalam memahami konteks bahasa alami, terutama dalam mendeteksi sentimen negatif dengan nilai F1-score tinggi (0,95). Meskipun demikian, performa model pada kategori sentimen positif dan netral masih relatif rendah, dengan F1-score masing-masing 0,52 dan 0,31. Hal ini menunjukkan adanya keterbatasan yang disebabkan oleh distribusi data yang tidak seimbang, di mana ulasan positif jauh lebih dominan dibandingkan kategori lainnya. Kondisi ini berdampak pada kesulitan model dalam membedakan sentimen positif dan netral. Oleh karena itu, penelitian selanjutnya perlu mempertimbangkan strategi penyeimbangan data, seperti oversampling, undersampling, atau penggunaan teknik data augmentation. Selain itu, pengujian dengan variasi model deep learning atau pendekatan hibrid juga berpotensi meningkatkan performa. Dengan perbaikan ini, analisis sentimen diharapkan dapat memberikan gambaran yang lebih akurat mengenai pengalaman dan kebutuhan pengguna aplikasi seni digital.

DAFTAR PUSTAKA

- Alaparthi, S., & Mishra, M. (2020). *Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey*. 1, 1-20. <https://doi.org/10.48550/arXiv.2007.01127>
- Albab, M. U., P., Y. K., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, 20(2), 1–12. <https://doi.org/10.26623/transformatika.v20i2.5374>
- Aliero, A. A., Adebayo, B. S., Aliyu, H. O., Tafida, A. G., Kangiwa, B. U., & Dankolo, N. M. (2023). Systematic Review on Text Normalization Techniques and its Approach to Non-Standard Words. *International Journal of Computer Applications*, 185(33), 44–55. <https://doi.org/10.5120/ijca2023923106>
- Areshey, A., & Mathkour, H. (2023). Transfer Learning For Sentiment Classification Using Bidirectional Encoder Representations From Transformers (BERT) Model. *Sensors*, 23(11). <https://doi.org/10.3390/s23115232>
- Cahyawijaya, S., Winata, G. I., Wilie, B., Vincentio, K., Li, X., Kuncoro, A., Ruder, S., Lim, Z. Y., Bahar, S., Khodra, M. L., Purwarianti, A., & Fung, P. (2021). IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 8875–8898. <https://doi.org/10.18653/v1/2021.emnlp-main.699>
- Diaz, G. (2025). *Stopwords-id*. <https://github.com/stopwords-iso/stopwords-id>
- Hadiprakoso, R. B., Setiawan, H., Yasa, R. N., & Girinoto. (2023). Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding. *AIP Conference Proceedings*, 2680(1). <https://doi.org/10.1063/5.0126116>
- Hasanudin, C., Mayasari, N., Saddhono, K., & Prabowo, R. A. (2021). IbisPaint X Apps in Creating Collaborative 3D Learning media of Pop-Up and Movable Books. *Journal of Physics: Conference Series*, 1764(1). <https://doi.org/10.1088/1742-6596/1764/1/012131>
- Jupin-Delevaux, É., Djahnine, A., Talbot, F., Richard, A., Gouttard, S., Mansuy, A., Douek, P., Si-Mohamed, S., & Boussel, L. (2023). BERT-Based Natural Language Processing Analysis Of French CT Reports: Application To The Measurement Of The Positivity Rate For Pulmonary Embolism. *Research in Diagnostic and Interventional Imaging*, 6, 100027. <https://doi.org/10.1016/j.redii.2023.100027>
- Latif, R. M. A., Abdullah, M. T., Shah, S. U. A., Farhan, M., Ijaz, F., & Karim, A. (2019). Data Scraping From Google Play Store And Visualization Of Its Content For Analytics. *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies, ICoMET 2019, July 2020*. <https://doi.org/10.1109/ICOMET.2019.8673523>
- Li, H., Xue, T., Zhang, A., Luo, X., Kong, L., & Huang, G. (2024). The Application And Impact Of Artificial Intelligence Technology In Graphic Design: A Critical Interpretive Synthesis. *Heliyon*, 10(21), e40037. <https://doi.org/10.1016/j.heliyon.2024.e40037>
- Maulana, A., Rozikin, C., & Surharso, A. (2025). Sentiment Analysis Of Indonesia National Team Naturalization Using Bidirectional Encoder Representations From Transformers. *Informatika Dan Sains*, 15(01), 2025. <https://doi.org/10.54209/infosains.v15i01>
- Mustak, M., Hallikainen, H., Laukkanen, T., Plé, L., Hollebeek, L. D., & Aleem, M. (2024). Using Machine Learning To Develop Customer Insights From User-Generated Content. *Journal of Retailing and Consumer Services*, 81(July). <https://doi.org/10.1016/j.jretconser.2024.104034>
- Nawawi, I., Ilmawan, K. F., Maarif, M. R., & Syafrudin, M. (2024). Exploring Tourist Experience through Online Reviews Using Aspect-Based Sentiment Analysis with Zero-Shot Learning for Hospitality

Service Enhancement. *Information (Switzerland)*, 15(8). <https://doi.org/10.3390/info15080499>

Oliaee, A. H., Das, S., Liu, J., & Rahman, M. A. (2023). Using Bidirectional Encoder Representations From Transformers (BERT) To Classify Traffic Crash Severity Types. *Natural Language Processing Journal*, 3, 100007. <https://doi.org/10.1016/j.nlp.2023.100007>

Pradipta, D., & Widodo, E. (2024). Sentiment Analysis on Social Media using Bidirectional Encoder from Transformers (Case Study : Covid – 19 Omicron). *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, 16(2), 267–281. <https://doi.org/10.37424/informasi.v16i2.319>

Rayhan, R., Rayhan, A., & Kinzler, R. (2023). Exploring the Power of Data Manipulation and Analysis: A Comprehensive Study of NumPy, SciPy, and Pandas. *ResearchGate*, 1-23. <https://doi.org/10.13140/RG.2.2.22390.16968>

Wang, J., & Zhang, J. (2024). Research on the Theory and Innovation Strategy of Digital Media Art Development Based on Artificial Intelligence Technology. *Applied Mathematics and Nonlinear Sciences*, 9(1), 1-20. <https://doi.org/10.2478/amns-2024-1543>