



Analisis Sentimen Ulasan iPhone di Amazon Menggunakan Model Deep Learning BERT Berbasis Transformer

Arif Fitra Setyawan¹, Amelia Devi Putri Ariyanto², Fari Katul Fikriah³, Rozaq Isnaini Nugraha⁴
ariffitra.setyawan@gmail.com¹, ameliadev26@gmail.com², farichatulfikriyah45@gmail.com³,
rozaqin@uwhs.ac.id⁴

Universitas Widya Husada

Sistem dan Teknologi Informasi

Jl. Subali Raya No.12, Krapyak, Kec. Semarang Barat, Kota Semarang, Jawa Tengah

ARTICLE INFO

Article history:

Received : 29 Agustus 2024

Received in revised : 18 November 2024

Accepted : 2 Desember 2024

Available online : 12 Desember 2024

ABSTRACT

This study aims to analyze the sentiment of iPhone product reviews from Amazon using the BERT (Bidirectional Encoder Representations from Transformers) model to classify reviews as either positive or negative. The dataset, sourced from Kaggle, includes text reviews and star ratings, where high ratings indicate positive sentiment and low ratings indicate negative sentiment. After text preprocessing steps, including data cleaning, tokenization, and sentiment labeling, the BERT model was fine-tuned for sentiment classification, with the data split into training, validation, and test sets. Evaluation results demonstrate that the BERT model achieves a high classification accuracy, with an accuracy rate of 93.9% and a balanced F1 score between precision and recall. Confusion matrix evaluation also indicates that the model consistently identifies both positive and negative sentiments. This study shows that Transformer-based models like BERT are highly effective in understanding customer opinions in e-commerce, with broad application potential for data-driven decision-making in marketing strategies and product development.

Keywords: Amazon, BERT, e-commerce, iPhone reviews, sentiment analysis, text classification, Transformer model

1. Pendahuluan

Kemajuan *e-commerce* telah secara signifikan mengubah pola belanja konsumen, namun tantangan berupa ketidaksesuaian antara kualitas produk yang sebenarnya dengan deskripsi yang disampaikan oleh penjual masih menjadi perhatian (Amelia, Fari, & Arif, 2024). Selain itu, perkembangan *e-commerce* juga berdampak pada meningkatnya peran ulasan produk online dalam memengaruhi proses pengambilan keputusan konsumen. Ulasan ini tidak hanya membantu konsumen lain dalam memilih produk, tetapi juga menyediakan umpan balik yang berharga bagi produsen untuk meningkatkan kualitas produk dan layanan mereka (Genoveva & Herny, 2024). Amazon.com Inc., sebagai perusahaan teknologi multinasional yang berpusat di Amerika Serikat, telah menjadi salah satu pelopor dalam industri *e-commerce global*. Perusahaan ini merupakan

bagian dari Amazon.com Inc, sebuah perusahaan teknologi multinasional asal Amerika Serikat yang memiliki fokus utama pada perdagangan elektronik atau *e-commerce* (Jesica, Agus, Mochammad, Yoseph, & Diah, 2024). Sebagai bagian dari Amazon.com Inc., perusahaan ini sangat memperhatikan ulasan konsumen, yang menjadi faktor utama dalam memengaruhi keputusan pembelian dengan menyediakan informasi tambahan tentang kualitas produk sekaligus merepresentasikan pengalaman nyata pengguna sebelumnya. Ulasan konsumen menjadi salah satu faktor utama yang mempengaruhi keputusan pembelian, karena tidak hanya memberikan informasi tambahan tentang kualitas produk, tetapi juga merepresentasikan pengalaman nyata dari pengguna sebelumnya. Opini-opini dalam ulasan ini sangat bernilai dan jika dianalisis dengan baik, dapat membantu perusahaan, calon pembeli, dan pemasar memahami persepsi publik terhadap produk yang ditawarkan. Analisis sentimen atau "opinion mining" merupakan metode yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini sebagai sentimen positif, negatif, atau netral (Muhammad, Ade, Rayhan, & Ridha, 2023). Analisis sentimen adalah metode yang digunakan untuk mengidentifikasi dan mengevaluasi opini, penilaian, serta emosi yang terdapat dalam sebuah teks (Diah, Bucky, & Dudi, 2024). Penggunaan analisis sentimen berbasis kecerdasan buatan memungkinkan proses identifikasi pola sentimen dari ulasan secara lebih efisien dan akurat dibandingkan analisis manual, terutama dengan memanfaatkan teknologi *Natural Language Processing* (NLP) yang terus berkembang. Analisis sentimen menjadi solusi yang banyak digunakan dalam pengolahan bahasa alami (*Natural Language Processing/NLP*) untuk mengidentifikasi opini pengguna dari teks ulasan (Yusuf & Tommy, 2020).

Kemajuan dalam NLP dan model *deep learning*, seperti BERT (*Bidirectional Encoder Representations from Transformers*), telah meningkatkan akurasi analisis sentimen karena kemampuannya memahami konteks teks secara bidirectional, melampaui model-model sebelumnya seperti LSTM dan RNN (Hadi, Muhammad, & Sigit, 2022). BERT, yang menggunakan pra-pelatihan pemodelan dua arah, dapat memahami konteks dari sisi kiri maupun kanan dalam teks, yang menjadikannya unggul dalam analisis sentimen (Jacob, Ming-Wei, Kenton, & Kristina, 2019). Dalam penelitian ini, model BERT berbasis Transformer digunakan untuk menganalisis sentimen ulasan produk iPhone di Amazon, dengan tujuan mengidentifikasi sentimen positif dan negatif serta mengevaluasi kinerja model menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*. Diharapkan hasil penelitian ini tidak hanya memberikan wawasan tentang persepsi konsumen terhadap produk iPhone, tetapi juga menjadi referensi untuk pengembangan lanjutan penggunaan analisis sentimen berbasis *Transformer* di *platform e-commerce* lainnya.

2. Metode penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis *Natural Language Processing* (NLP) dengan model BERT (*Bidirectional Encoder Representations from Transformers*) untuk klasifikasi sentimen pada ulasan iPhone dari platform Amazon. Model BERT menghasilkan representasi kontekstual dari teks input melalui proses tokenisasi menjadi token kecil, *embedding* untuk mengkonversi token menjadi vektor kata, serta pengolahan oleh *encoder* bertingkat yang mengandalkan *multi-head self-attention* dan *feedforward networks* untuk menangkap konteks setiap token dalam teks (Andi & Binti, 2024). BERT berperan dalam tokenisasi, *embedding*, dan klasifikasi teks untuk mengidentifikasi pola dan karakteristik dalam teks (Asep, Firman, & Farihin, 2024). Tahapan metode BERT meliputi Pengumpulan Data, Pra-pemrosesan Data, Pembagian Data, Pelatihan dan *Fine-tuning* Model BERT, Evaluasi Model menggunakan metrik, serta Analisis Hasil Evaluasi.

2.1. Pengumpulan Data

Data ulasan produk iPhone dari Amazon diambil dari Kaggle dalam format CSV dan diproses untuk analisis sentimen. Dataset berisi teks ulasan, rating bintang dan beberapa metadata tambahan. Label sentimen ditentukan berdasarkan rating, dengan ulasan bernilai 4 atau 5 dikategorikan sebagai sentimen positif, dan rating 1, 2, atau 3 sebagai sentimen negatif. Data yang kosong, duplikat, atau tidak lengkap dihapus, sementara distribusi rating ditinjau untuk menjaga keseimbangan antara ulasan positif dan negatif. Data yang telah dibersihkan disimpan dan siap untuk digunakan pada tahap pra-pemrosesan teks dan pelatihan model BERT berbasis Transformer.

2.2. Pra-pemrosesan Data

Setelah data terkumpul, data ulasan teks diproses lebih lanjut agar dapat digunakan sebagai input model secara efektif. Tahapan pra-pemrosesan yang dilakukan meliputi beberapa langkah penting. Pertama, tokenisasi dilakukan dengan memecah teks menjadi unit-unit token menggunakan tokenizer khusus untuk BERT, seperti *BertTokenizer* dari pustaka *Transformers*. Langkah ini memastikan teks memiliki format yang sesuai dengan kebutuhan model BERT, termasuk penambahan token khusus seperti [CLS] dan [SEP]. Selanjutnya, semua huruf dalam teks diubah menjadi huruf kecil (*lowercasing*) untuk menjaga konsistensi format dan menghindari perbedaan akibat variasi kapitalisasi. Setelah itu, dilakukan pembersihan teks dengan menghapus tanda baca, angka, simbol, dan karakter khusus lainnya yang tidak relevan untuk analisis sentimen. Proses ini diikuti dengan penghapusan kata-kata umum (*stop word removal*), seperti "the", "and" dan "is", untuk mengurangi kebisingan data. Terakhir, teks hasil tokenisasi dipotong atau diisi (*padding*) hingga mencapai panjang maksimum yang ditentukan model, biasanya 128 token, untuk memastikan konsistensi panjang input. Tahapan ini memastikan bahwa data siap digunakan dalam analisis berbasis model BERT secara optimal.

2.3. Pembagian Data

Setelah proses pra-pemrosesan selesai, data dibagi menjadi tiga subset utama dengan proporsi sebagai berikut. Pertama, data pelatihan (*train set*) yang mencakup 70% dari keseluruhan data digunakan untuk melatih model. Kedua, data validasi (*validation set*) yang mencakup 15% dari data dimanfaatkan untuk mengevaluasi kinerja model selama proses pelatihan, mencegah terjadinya *overfitting*, serta menyesuaikan parameter hiper (*hyperparameter*). Ketiga, data uji (*test set*) yang mencakup 15% dari data digunakan untuk evaluasi akhir guna mengukur kemampuan model pada data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk memastikan pengujian model dilakukan secara terstruktur dan menyeluruh.

2.4. Pelatihan dan Fine-tuning Model BERT

Proses pelatihan model dilakukan dengan metode *fine-tuning* pada model BERT, yang sebelumnya telah dilatih secara umum menggunakan korpus teks dalam bahasa Inggris. Model BERT diinstansiasi dengan memanfaatkan pustaka *Transformers* dari *Hugging Face*, menggunakan varian dasar seperti *bert-base-uncased*.

Pada tahap *fine-tuning*, model dilengkapi dengan penambahan layer klasifikasi pada lapisan akhir BERT, yang dirancang untuk mengklasifikasikan keluaran menjadi dua kategori sentimen, yaitu positif atau negatif. Proses pelatihan dilakukan menggunakan *optimizer* dengan penyesuaian *learning rate* melalui *scheduler* yang dirancang untuk meningkatkan efisiensi dan efektivitas pelatihan.

Parameter pelatihan seperti *batch size* ditetapkan sebesar 16 atau 32, sedangkan jumlah *epoch* disesuaikan antara 3 hingga 5, tergantung pada tingkat konvergensi model berdasarkan evaluasi pada data validasi. Untuk mencegah *overfitting*, diterapkan teknik regularisasi seperti *dropout* pada

lapisan akhir serta *weight decay* untuk menjaga *generalisasi* model. Proses ini bertujuan untuk mengoptimalkan kinerja model pada tugas analisis sentimen.

2.5. Evaluasi Model dengan Metrik

Untuk mengevaluasi performa model, diterapkan berbagai metrik evaluasi pada data uji untuk memberikan penilaian yang komprehensif terhadap kemampuan model. Akurasi digunakan untuk menghitung persentase prediksi yang benar dibandingkan dengan seluruh data uji. Presisi mengukur persentase prediksi positif yang benar-benar positif, yang bertujuan untuk mengurangi kesalahan berupa prediksi positif palsu (*false positive*). *Recall* menilai persentase data positif yang berhasil diidentifikasi dengan benar oleh model, dengan fokus pada pengurangan kesalahan berupa prediksi negatif palsu (*false negative*).

Untuk mendapatkan evaluasi yang seimbang antara presisi dan recall, digunakan *F1 Score*, yaitu rata-rata harmonis dari kedua metrik tersebut, yang sangat penting terutama pada dataset dengan distribusi kelas yang tidak seimbang. Selain itu, *confusion matrix* digunakan untuk memvisualisasikan hasil klasifikasi, baik klasifikasi yang benar maupun kesalahan antar kelas sentimen. Matriks ini memberikan informasi terperinci mengenai kemampuan model dalam mengenali setiap kelas, serta divisualisasikan untuk mempermudah interpretasi. Secara keseluruhan, metrik-metrik ini dirancang untuk memastikan evaluasi performa model dilakukan secara holistik dan mendalam.

3. Hasil dan Pembahasan

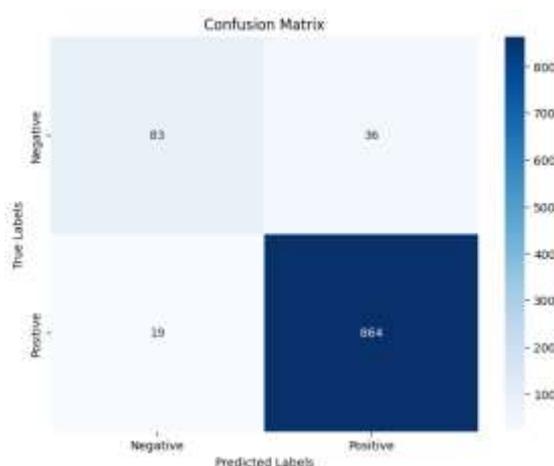
Model BERT yang digunakan dalam penelitian ini menunjukkan performa yang sangat baik dalam tugas klasifikasi sentimen terhadap ulasan produk iPhone di platform Amazon. Hasil evaluasi menunjukkan tingkat akurasi dan F1 Score yang tinggi, mengindikasikan kemampuan model dalam mengidentifikasi sentimen ulasan secara efektif. Selain itu, keseimbangan antara nilai *precision* dan *recall* mencerminkan performa model yang konsisten dalam mengklasifikasikan sentimen dengan benar, baik dalam mendeteksi ulasan positif maupun negatif. Hal ini menegaskan bahwa model memiliki tingkat generalisasi yang baik pada data uji. Tabel berikut menyajikan hasil pengujian model BERT;

Tabel 1. hasil pengujian model BERT

Epoch	Training Loss	Validation Loss	Accuracy
1	No log	0.199373	0.930140
2	0.192700	0.181465	0.945110
3	0.192700	0.189479	0.945110

Selama proses pelatihan yang berlangsung selama tiga *epoch*, model menunjukkan penurunan *validation loss* yang signifikan pada *epoch* pertama, dari 0.1993 menjadi 0.1815 pada *epoch* kedua. Meskipun pada *epoch* ketiga *validation loss* mengalami sedikit peningkatan menjadi 0.1895, akurasi model tetap stabil dan konsisten pada nilai 94,51%.

Sebagai bagian dari evaluasi performa, *confusion matrix* digunakan untuk menganalisis jumlah kesalahan prediksi antara kelas positif dan negatif. Analisis ini membantu dalam mengidentifikasi area yang memerlukan perbaikan lebih lanjut, sehingga dapat meningkatkan kemampuan model dalam mengklasifikasikan sentimen secara akurat. Hasil *confusion matrix* disajikan pada tabel berikut;



Gambar 1. Confusion Matrix

Dari hasil evaluasi menggunakan data uji, model mencapai akurasi sebesar 94,51%. Untuk menilai kinerja model lebih lanjut, digunakan metrik tambahan seperti precision, recall, dan F1 score. Nilai *precision* yang diperoleh sebesar 96,00% menunjukkan bahwa model memiliki tingkat akurasi yang tinggi dalam mengidentifikasi ulasan positif. Selain itu, nilai *recall* yang mencapai 97,85% mengindikasikan bahwa hampir seluruh ulasan positif berhasil terdeteksi dengan baik oleh model. Hasil ini semakin diperkuat dengan *F1 score* sebesar 96,92%, yang mencerminkan keseimbangan optimal antara *precision* dan *recall*.

Secara keseluruhan, performa model yang tinggi pada berbagai metrik evaluasi menunjukkan bahwa model BERT mampu mengenali dan mengklasifikasikan sentimen ulasan pengguna dengan efektif. Dengan akurasi yang hampir mencapai 95%, model ini terbukti efisien dalam melakukan analisis sentimen untuk memahami persepsi pengguna terhadap produk iPhone di Amazon. Hasil ini juga menunjukkan potensi besar penerapan model berbasis Transformer dalam analisis sentimen di platform *e-commerce*, yang dapat memberikan manfaat signifikan bagi perusahaan dalam mengambil keputusan berbasis data untuk meningkatkan strategi pemasaran dan pengembangan produk.

4. Conclusion

Penelitian ini berhasil menerapkan model BERT berbasis Transformer untuk menganalisis sentimen pada ulasan produk iPhone di Amazon. Dengan memanfaatkan arsitektur Transformer yang memungkinkan pemahaman konteks secara menyeluruh, model ini mampu mengidentifikasi sentimen positif dan negatif dengan akurasi yang tinggi. Hasil evaluasi menunjukkan performa yang sangat baik, yang tercermin dalam skor akurasi, presisi, *recall* dan F1 yang memuaskan, serta distribusi yang seimbang pada *confusion matrix*. Hal ini menunjukkan bahwa model BERT mampu mengenali pola bahasa dalam teks ulasan dengan baik, menghasilkan klasifikasi sentimen yang akurat.

Penerapan model BERT pada analisis sentimen ulasan produk memberikan manfaat besar bagi bisnis *e-commerce* dan produsen, terutama dalam memahami persepsi konsumen secara cepat dan efektif. Informasi yang diperoleh dari analisis sentimen ini dapat membantu perusahaan dalam mengambil keputusan yang lebih tepat terkait strategi pemasaran dan pengembangan produk, berdasarkan respons aktual dari pengguna. Lebih lanjut, model ini memiliki potensi untuk diterapkan pada produk atau platform lain, menjadikannya solusi yang fleksibel untuk berbagai keperluan analisis sentimen di *e-commerce*.

Meskipun hasil yang diperoleh menjanjikan, terdapat beberapa area yang dapat dikembangkan lebih lanjut. Penggunaan model dengan kapasitas yang lebih besar, seperti RoBERTa atau GPT

dapat menjadi alternatif untuk meningkatkan performa model. Selain itu, eksplorasi teknik *fine-tuning* yang lebih kompleks serta penambahan data ulasan yang lebih beragam dapat memberikan hasil yang lebih akurat dan dapat digeneralisasi. Ke depan, penelitian ini membuka peluang untuk pengembangan model analisis sentimen yang lebih mendalam dan adaptif, yang dapat memenuhi berbagai kebutuhan analisis teks dalam lingkungan digital yang terus berkembang.

References

- Amelia, D. P., Fari, K. F., & Arif, F. S. (2024). Emotion Detection Using Contextual Embeddings for Indonesian Product Review Texts on E-commerce Platform. *Pixel: Jurnal Ilmiah Komputer Grafis*, 179-185.
- Andi, A., & Binti, M. K. (2024). Mengungkap Opini Publik: Pendekatan BERT-based untuk Analisis Sentimen pada Komentar Film. *Journal of System and Computer Engineering (JSCE)*, 36-43.
- Asep, R., Firman, S., & Farihin, L. (2024). Deteksi Berita Hoax dengan Perbandingan Website Menggunakan Pendekatan Deep Learning Algoritma BERT. *G-Tech : Jurnal Teknologi Terapan*, 1749-1758.
- Diah, F. S., Bucky, W. K., & Dudi, Y. (2024). Analisis Sentimen Film Dirty Vote Menggunakan BERT (Bidirectional Encoder Representations from Transformers). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 393-404.
- Genoveva, A., & Herny, F. (2024). Analisis Sentimen Ulasan Herborist Sistem Pengambilan Keputusan Menggunakan Klasifikasi Neighbordan TF-ID. *JURNAL SWABUMI*, 176-181.
- Hadi, P., Muhammad, F., & Sigit, W. (2022). Implementasi BERT pada Analisis Sentimen di Media Sosial. *Jurnal Informatika dan Sistem Informasi*, 45-54.
- Jacob, D., Ming-Wei, C., Kenton, L., & Kristina, T. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*.
- Jesica, Y. B., Agus, J., Mochammad, A. A., Yoseph, T. A., & Diah, P. (2024). Analisis Sentimen Produk Amazon Menggunakan Algoritma Naïve Bayes Pada Data Review Pelanggan. *Jurnal Teknik Komputer AMIK BSI*, 13-19.
- Muhammad, I., Ade, D. W., Rayhan, S., & Ridha, F. A. (2023). Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 1799-1807.
- Yusuf, S., & Tommy, W. (2020). Penggunaan NLP untuk Analisis Sentimen Produk Smartphone. *Jurnal Teknologi dan Informasi*, 200-210.